**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC1/SC29/WG11**

**CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC1/SC29/WG11 N18028**

**October 2018, Macao, CN**

|  |  |
| --- | --- |
| **Source** | **JVET** |
| **Status** | **Approved** |
| **Title** | **Test Model 3 of Versatile Video Coding (VTM 3)** |

|  |  |
| --- | --- |
| Joint Video Experts Team (JVET)  of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11  12th Meeting: Macao, CN, 3–12 Oct. 2018 | Document: JVET-L1002-v1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Algorithm description for Versatile Video Coding and Test Model 3 (VTM 3)** | | |
| *Status:* | Output document of JVET | | |
| *Purpose:* | Algorithm description for Versatile Video Coding and Test Model 3 | | |
| *Author(s) or Contact(s):* | Jianle Chen Yan Ye Seung Hwan Kim | Email: | [jianle.chen@huawei.com](mailto:jianle.chen@huawei.com) [Yan.Ye@alibaba-inc.com](mailto:Yan.Ye@alibaba-inc.com) seunghwan3.kim@lge.com |
| *Source:* | Editors | | |

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

# Abstract

The JVET established the Versatile Video Coding (VVC) working draft 3 and the VVC Test Model 3 (VTM3) algorithm description and encoding method at its 12th meeting (3–12 October, 2018, Macao, CN). This document serves as a source of general tutorial information on the VVC design and also provides an encoder-side description of VTM3.

Ed. Notes:

VVC Test Model 3 (VTM3) algorithm description and encoding method

* Incorporated Adaptive Loop Filter
  + JVET-L0082: 10 b coeffs (instead of 11)
  + JVET-L0147: Subsampled Laplacian calculation
  + JVET-L0083: Reduction of bits for ALF coefficient fractional part
  + JVET-L0392: minor BF
  + JVET-L0664: Remove the signaling of 5x5 as a special case for luma
* JVET-L0081: 64x64 luma size virtual pipeline data units (VPDUs)
* Incorporated Affine related modification, including
  + JVET-L0265: set the chroma subblock size to 4x4 instead of 2x2
  + JVET-L0271: CE4.1.6: Simplification of affine AMVP candidate list construction
  + JVET-L0045: line buffer reduction for affine mode
  + JVET-L0632/L0142: affine merge refinement
  + JVET-L0369/L0055 : moving ATMVP into the affine merge list
* JVET-L0293: CPR mode for screen content coding
* JVET-L0646: bi-prediction with weighted averaging
* JVET-L0256: bi-directional optical flow
* JVET-L0231: horizontal wrap-around motion compensation
* JVET-L0377: Rounding Align of Adaptive Motion Vector Resolution
* JVET-L0198/L0468/L0104: fixed subblock size of 8x8 for SbTMVP mode
* JVET-L0104: disallow 4x4 bi-prediction
* Incorporated JVET-L0191: CCLM parameter derivation
* Incorporated JVET-L0136/JVET-L0085: CCLM with line buffer restriction
* Incorporated JVET-L0338/JVET-L0340: Multi-directional LM (MDLM)
* Incorporated JVET-L0053/JVET-L0272: chroma DM based on center position
* Incorporated JVET-L0279: unification of angular intra prediction
* Incorporated JVET-L0165: intra 6 MPM
* Incorporated JVET-L0059: simplification on MTS kernel derivation
* Incorporated JVET-L0111: transform skip condition on transform block size
* Incorporated JVET-L0285: 8-bit transform matrices
* Incorporated JVET-L0118: unified MTS signaling
* Incorporated JVET-L0553: quantization semantics fix
* Incorporated JVET-L0274: coefficient coding
* Incorporated JVET-L0628: mode dependent intra smoothing
* Incorporated JVET-L0283: multiple reference line intra prediction
* Incorporated JVET-L0414: DF strength dependent on reconstructed luma level
* Incorporated JVET-L0410: Deblocking tC table
* JVET\_L0124/L0208: triangle partition mode
* JVET-L0100: combined intra and inter prediction
* Added merge list generation process, including
  + Spatial MVP and Temporal MVP derivation
  + JVET-L0266/: History-based MVP from an FIFO table
  + JVET-L0090: Pairwise average MVP
* Incorporated JVET-L0054: merge with MVD (MMVD)

VVC Test Model 2 (VTM2) algorithm description and encoding method

* Incorporated JVET-K0230: Separate trees for intra slices (without multi-DMs) with an implicit split to 64x64;
* Incorporated JVET-K0556: Prohibit ternary split of something bigger than 64 in width or height (and not send the bit to indicate ternary type at that level).
* Incorporated JVET-K0351 (test c): Keep only the TT restriction (preventing binary split with same orientation in center partition of the ternary split)
* Incorporated JVET-K0554: Implicit splitting at picture boundaries and ensure MinQTSize at boundary splits
* Incorporated JVET-K0063: Position dependent intra prediction combination (PDPC)
* Incorporated JVET-K0190: CCLM only (test 4.1.8)
* Incorporated JVET-K0122: DC prediction bug fix
* Incorporated JVET-K0529: 67 modes with 3MPM and FLC for non-MPM
* Incorporated JVET-K0500: Wide-angle intra prediction for non-square block
* Incorporated MTS (AMT) modification: Multiple transform selection (MTS)
* Incorporated sub-block TMVP
* Incorporated adaptive motion vector resolution
* Incorporated 8x8 and 1/16 pel motion field storage
* Incorporated affine motion

# Introduction

At the 10th JVET meeting (April 10–20, 2018, San Diego, US), JVET defined the first draft of Versatile Video Coding (VVC) and the VVC Test Model 1 (VTM1) encoding method. It was decided to include a quadtree with nested multi-type tree using binary and ternary splits coding block structure as the initial new coding feature of VVC. Draft reference software to implement the VTM1 encoding method (and the draft VVC decoding process) has also been developed. At the 11th meeting (10–18 July, 2018, Ljubljana, SI), the Versatile Video Coding (VVC) working draft 2 and the VVC Test Model 2 (VTM2) algorithm description and encoding method were established with the inclusion of a group of new coding features as well as some of HEVC coding elements. At the 12th meeting (3–12 October, 2018, Macao, CN), the Versatile Video Coding (VVC) working draft 3 and the VVC Test Model 3 (VTM3) algorithm description and encoding method were established with the inclusion of additional coding tools.

# Scope

The normative decoding process for Versatile Video Coding is specified in the VVC draft 3 text specification document [1]. The VTM3 reference software is provided to demonstrate a reference implementation of non-normative encoding techniques and the normative decoding process for VVC. The reference software can be accessed via

https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware\_VTM.git

This document provides an algorithm description as well as an encoder-side description of the VVC Test Model 3, which serves as a tutorial for the algorithm and encoding model implemented in the VTM3.0 software. The purpose of this document is to share a common understanding of the coding features of VVC and the reference encoding methods supported in the VTM3.0 software, in order to facilitate the assessment of the technical impact of new technologies during the standardization process. Common test conditions and software reference configurations that should be used for experimental work for conventional standard-dynamic range rectangular video content are described in JVET-L1010 [2]. Common test conditions specific to video content with high dynamic range and wide colour gamut are described in JVET-L1011 [3]. Common test conditions specific to video content for 360° omnidirectional video applications are described in JVET-L1012 [4]. When encoding and decoding 360° omnidirectional video, an additional software package called the 360Lib needs to be used together with using the VTM3.0 software to process, encode/decode and compute the spherical quality metrics. The 360Lib software is available at:

https://jvet.hhi.fraunhofer.de/svn/svn\_360Lib/

Additionally, document JVET-L1004 [5] describes the algorithms used in 360Lib to process, code, and measure quality of 360° omnidirectional video.

# Algorithm description of Versatile Video Coding

## VVC coding architecture

As in most preceding standards, VVC has a block-based hybrid coding architecture, combining inter-picture and intra-picture prediction and transform coding with entropy coding. Figure 1 shows a general block diagram of the VTM3 encoder.

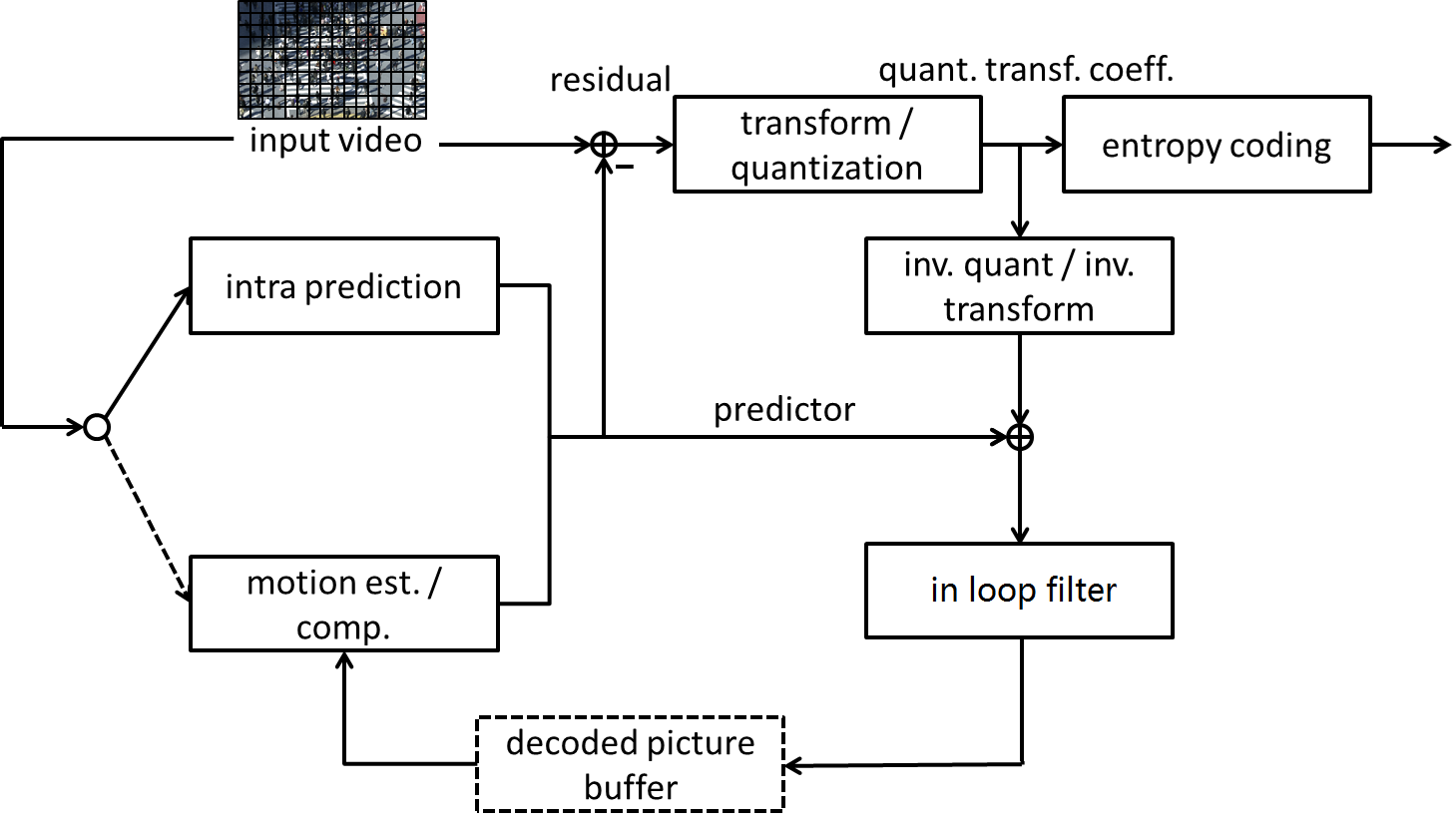


Figure 1 – General block diagram of VTM3 encoder

The picture partitioning structure, which is further described in section 3.2, divides the input video into blocks called coding tree units (CTUs). A CTU is split using a quadtree with nested multi-type tree structure into coding units (CUs), with a leaf coding unit (CU) defining a region sharing the same prediction mode (e.g. intra or inter). In this document, the term ‘unit’ defines a region of an image covering all colour components; the term ‘block’ is used to define a region covering a particular colour component (e.g. luma), and may differ in spatial location when considering the chroma sampling format such as 4:2:0.

The other features of VTM3, including intra prediction processes, inter picture prediction processes, transform and quantization processes, entropy coding processes and in-loop filter processes, are covered in sections 3.3 to 3.7. As agreed in the 11th JVET meeting, the following features have been included in the VVC test model 3 on top of the bock tree structure.

* Intra prediction
  + 67 intra mode with wide angles mode extension
  + Block size and mode dependent 4 tap interpolation filter
  + Position dependent intra prediction combination (PDPC)
  + Cross component linear model intra prediction
  + Multi-reference line intra prediction
* Inter-picture prediction
  + Block motion copy with spatial, temporal, history-based, and pairwise average merging candidates
  + Affine motion inter prediction
  + sub-block based temporal motion vector prediction
  + Adaptive motion vector resolution
  + 8x8 block based motion compression for temporal motion prediction
  + High precision (1/16 pel) motion vector storage and motion compensation with 8-tap interpolation filter for luma component and 4-tap interpolation filter for chroma component
  + Triangular partitions
  + Combined intra and inter prediction
  + Merge with MVD (MMVD)
  + Bi-directional optical flow
  + Bi-predictive weighted averaging
* Transform, quantization and coefficients coding
  + Multiple primary transform selection with DCT2, DST7 and DCT8
  + Dependent quantization with max QP increased from 51 to 63
  + Transform coefficient coding with sign data hiding
* In loop filter
  + Deblocking filter
  + Sample adaptive offset
  + Adaptive Loop Filter
* Screen content coding:
  + Current picture referencing with current CTU restriction
* 360-degree video coding
  + Horizontal wrap-around motion compensation
* Simple high-level syntax

## Partitioning

### Partitioning of the picture into CTUs

Pictures are divided into a sequence of coding tree units (CTUs). The CTU concept is same to that of the HEVC [6][7]. For a picture that has three sample arrays, a CTU consists of an N×N block of luma samples together with two corresponding blocks of chroma samples. Figure 1 shows the example of a picture divided into CTUs.

The maximum allowed size of the luma block in a CTU is specified to be 128×128 (although the maximum size of the luma transform blocks is 64×64).



Figure 2 – Example of a picture divided into CTUs

### Partitioning of the CTUs using a tree structure

In HEVC, a CTU is split into CUs by using a quaternary-tree structure denoted as coding tree to adapt to various local characteristics. The decision whether to code a picture area using inter-picture (temporal) or intra-picture (spatial) prediction is made at the leaf CU level. Each leaf CU can be further split into one, two or four PUs according to the PU splitting type. Inside one PU, the same prediction process is applied and the relevant information is transmitted to the decoder on a PU basis. After obtaining the residual block by applying the prediction process based on the PU splitting type, a leaf CU can be partitioned into transform units (TUs) according to another quaternary-tree structure similar to the coding tree for the CU. One of key feature of the HEVC structure is that it has the multiple partition conceptions including CU, PU, and TU.

In VVC, a quadtree with nested multi-type tree using binary and ternary splits segmentation structure replaces the concepts of multiple partition unit types, i.e. it removes the separation of the CU, PU and TU concepts except as needed for CUs that have a size too large for the maximum transform length, and supports more flexibility for CU partition shapes. In the coding tree structure, a CU can have either a square or rectangular shape. A coding tree unit (CTU) is first partitioned by a quaternary tree (a.k.a. quadtree) structure. Then the quaternary tree leaf nodes can be further partitioned by a multi-type tree structure. As shown in Figure 3, there are four splitting types in multi-type tree structure, vertical binary splitting (SPLIT\_BT\_VER), horizontal binary splitting (SPLIT\_BT\_HOR), vertical ternary splitting (SPLIT\_TT\_VER), and horizontal ternary splitting (SPLIT\_TT\_HOR). The multi-type tree leaf nodes are called coding units (CUs), and unless the CU is too large for the maximum transform length, this segmentation is used for prediction and transform processing without any further partitioning. This means that, in most cases, the CU, PU and TU have the same block size in the quadtree with nested multi-type tree coding block structure. The exception occurs when maximum supported transform length is smaller than the width or height of the colour component of the CU.



Figure 3 – Multi-type tree splitting modes

Figure 4 illustrates the signalling mechanism of the partition splitting information in quadtree with nested multi-type tree coding tree structure. A coding tree unit (CTU) is treated as the root of a quaternary tree and is first partitioned by a quaternary tree structure. Each quaternary tree leaf node (when sufficiently large to allow it) is then further partitioned by a multi-type tree structure. In the multi-type tree structure, a first flag (mtt\_split\_cu\_flag) is signalled to indicate whether the node is further partitioned; when a node is further partitioned, a second flag (mtt\_split\_cu\_vertical\_flag) is signalled to indicate the splitting direction, and then a third flag (mtt\_split\_cu\_binary\_flag) is signalled to indicate whether the split is a binary split or a ternary split. Based on the values of mtt\_split\_cu\_vertical\_flag and mtt\_split\_cu\_binary\_flag, the multi-type tree slitting mode (MttSplitMode) of a CU is derived as shown in Table 3‑1.

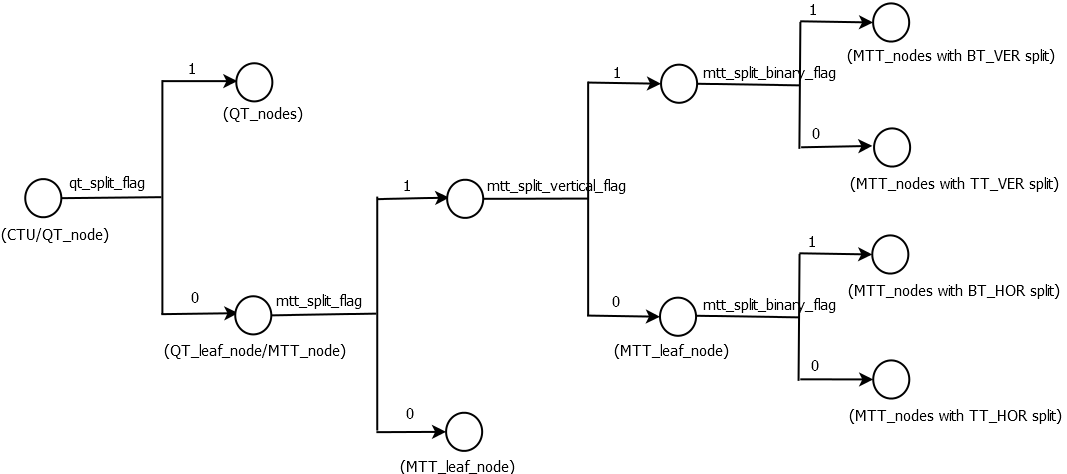


Figure 4 – Splitting flags signalling in quadtree with nested multi-type tree coding tree structure

**Table 3‑1 – MttSplitMode derviation based on multi-type tree syntax elements**

|  |  |  |
| --- | --- | --- |
| **MttSplitMode** | **mtt\_split\_cu\_vertical\_flag** | **mtt\_split\_cu\_binary\_flag** |
| SPLIT\_TT\_HOR | 0 | 0 |
| SPLIT\_BT\_HOR | 0 | 1 |
| SPLIT\_TT\_VER | 1 | 0 |
| SPLIT\_BT\_VER | 1 | 1 |

Figure 5 shows a CTU divided into multiple CUs with a quadtree and nested multi-type tree coding block structure, where the bold block edges represent quadtree partitioning and the remaining edges represent multi-type tree partitioning. The quadtree with nested multi-type tree partition provides a content-adaptive coding tree structure comprised of CUs. The size of the CU may be as large as the CTU or as small as 4×4 in units of luma samples. For the case of the 4:2:0 chroma format, the maximum chroma CB size is 64×64 and the minimum chroma CB size is 2×2.

In VVC, the maximum supported luma transform size is 64×64 and the maximum supported chroma transform size is 32×32. When the width or height of the CB is larger the maximum transform width or height, the CB is automatically split in the horizontal and/or vertical direction to meet the transform size restriction in that direction.

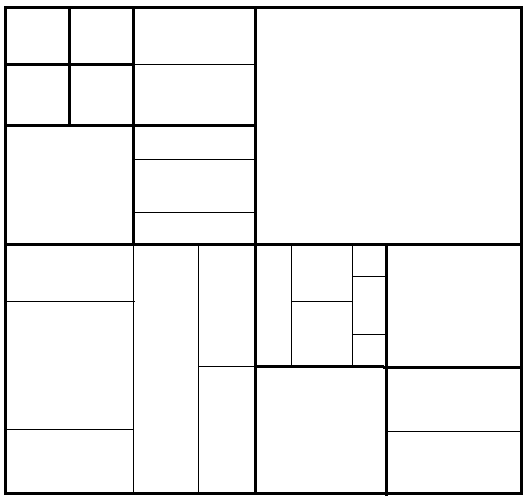


Figure 5– Example of quadtree with nested multi-type tree coding block structure

The following parameters are defined and specified by SPS syntax elements for the quadtree with nested multi-type tree coding tree scheme.

– CTU size: the root node size of a quaternary tree

– *MinQTSize*: the minimum allowed quaternary tree leaf node size

– *MaxBtSize*: the maximum allowed binary tree root node size

– *MaxTtSize*: the maximum allowed ternary tree root node size

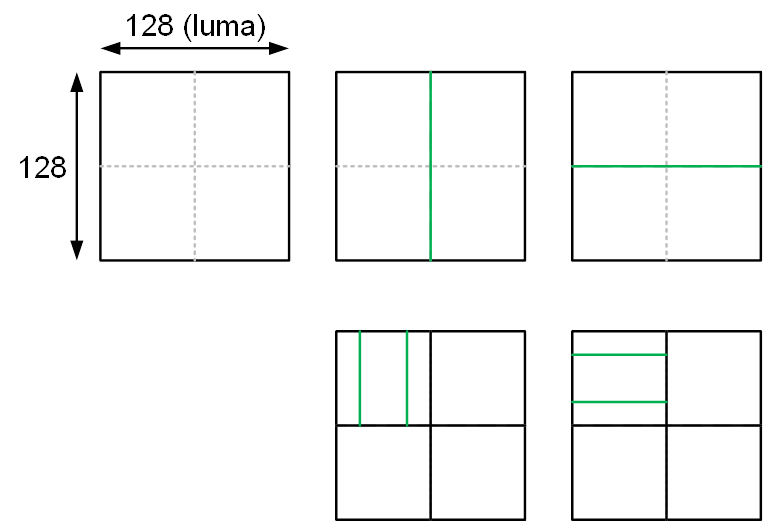
– *MaxMttDepth*: the maximum allowed hierarchy depth of multi-type tree splitting from a quadtree leaf

– *MinBtSize*: the minimum allowed binary tree leaf node size

– *MinTtSize*: the minimum allowed ternary tree leaf node size

In one example of the quadtree with nested multi-type tree coding tree structure, the CTU size is set as 128×128 luma samples with two corresponding 64×64 blocks of 4:2:0 chroma samples, the *MinQTSize* is set as 16×16, the *MaxBtSize* is set as 128×128and *MaxTtSize* is set as 64×64, the *MinBtSize* and *MinTtSize* (for both width and height) is set as 4×4, and the *MaxMttDepth* is set as 4. The quaternary tree partitioning is applied to the CTU first to generate quaternary tree leaf nodes. The quaternary tree leaf nodes may have a size from 16×16 (i.e., the *MinQTSize*) to 128×128 (i.e., the CTU size). If the leaf QT node is 128×128, it will not be further split by the binary tree since the size exceeds the *MaxBtSize* and *MaxTtSize* (i.e., 64×64). Otherwise, the leaf qdtree node could be further partitioned by the multi-type tree. Therefore, the quaternary tree leaf node is also the root node for the multi-type tree and it has multi-type tree depth (mttDepth) as 0. When the multi-type tree depth reaches *MaxMttDepth* (i.e., 4), no further splitting is considered. When the multi-type tree node has width equal to *MinBtSize* and smaller or equal to 2 \* *MinTtSize*, no further horizontal splitting is considered. Similarly, when the multi-type tree node has height equal to *MinBtSize and* smaller or equal to 2 \* *MinTtSize*, no further vertical splitting is considered.

To allow 64×64 Luma block and 32×32 Chroma pipelining design in VVC hardware decoders, TT split is forbidden when either width or height of a luma coding block is larger than 64 , as shown in Figure 6. TT split is also forbidden when either width or height of a chroma coding block is larger than 32.



**Figure 6– No TT split for 128×128 coding block**

In VTM3, the coding tree scheme supports the ability for the luma and chroma to have a separate block tree structure. Currently, for P and B slices, the luma and chroma CTBs in one CTU have to share the same coding tree structure. However, for I slices, the luma and chroma can have separate block tree structures. When separate block tree mode is applied, luma CTB is partitioned into CUs by one coding tree structure, and the chroma CTBs are partitioned into chroma CUs by another coding tree structure. This means that a CU in an I slice may consist of a coding block of the luma component or coding blocks of two chroma components, and a CU in a P or B slice always consists of coding blocks of all three colour components unless the video is monochrome.

### CU splits on picture boundaries

As done in HEVC, when a portion of a tree node block exceeds the bottom or right picture boundary, the tree node block is forced to be split until the all samples of every coded CU are located inside the picture boundaries. The following splitting rules are applied in the VTM3:

– If a portion of a tree node block exceeds both the bottom and the right picture boundaries,

* + If the block is a QT node and the size of the block is larger than the minimum QT size, the block is forced to be split with QT split mode.
  + Otherwise, the block is forced to be split with SPLIT\_BT\_HOR mode

– Otherwise if a portion of a tree node block exceeds the bottom picture boundaries,

* + If the block is a QT node, and the size of the block is larger than the minimum QT size, and the size of the block is larger than the maximum BT size, the block is forced to be split with QT split mode.
  + Otherwise, if the block is a QT node, and the size of the block is larger than the minimum QT size and the size of the block is smaller than or equal to the maximum BT size, the block is forced to be split with QT split mode or SPLIT\_BT\_HOR mode.
  + Otherwise (the block is a BTT node or the size of the block is smaller than or equal to the minimum QT size), the block is forced to be split with SPLIT\_BT\_HOR mode.

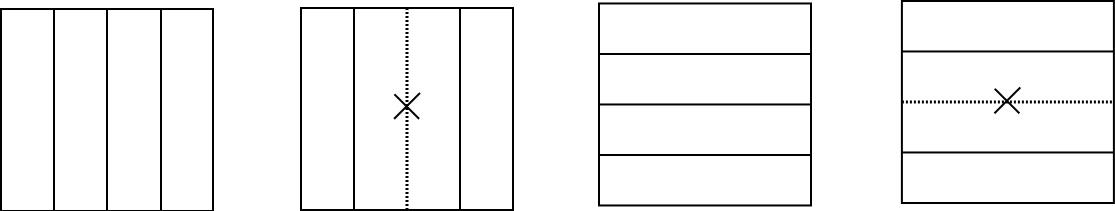
– Otherwise if a portion of a tree node block exceeds the right picture boundaries,

* + If the block is a QT node, and the size of the block is larger than the minimum QT size, and the size of the block is larger than the maximum BT size, the block is forced to be split with QT split mode.
  + Otherwise, if the block is a QT node, and the size of the block is larger than the minimum QT size and the size of the block is smaller than or equal to the maximum BT size, the block is forced to be split with QT split mode or SPLIT\_BT\_VER mode.
  + Otherwise (the block is a BTT node or the size of the block is smaller than or equal to the minimum QT size), the block is forced to be split with SPLIT\_BT\_VER mode.

### Restrictions on redundant CU splits

The quadtree with nested multi-type tree coding block structure provides a highly flexible block partitioning structure. Due to the types of splits supported the multi-type tree, different splitting patterns could potentially result in the same coding block structure. In VVC, some of these redundant splitting patterns are disallowed.

Figure 7 illustrates the redundant splitting patterns of binary tree splits and ternary tree splits. As shown in Figure 7, two levels of consecutive binary splits in one direction could have the same coding block structure as a ternary tree split followed by a binary tree split of the central partition. In this case, the binary tree split (in the given direction) for the central partition of a ternary tree split is prevented by the syntax. This restriction applies for CUs in all pictures.



**Figure 7–Redundant splitting patterns of binary tree split and ternary tree split cases**

When the splits are prohibited as described above, signalling of the corresponding syntax elements is modified to account for the prohibited cases. For example, when any case in Figure 7 is identified (i.e. the binary split is prohibited for a CU of a central partition), the syntax element mtt\_split\_cu\_binary\_flag which specifies whether the split is a binary split or a ternary split is not signalled and is instead inferred to be equal to 0 by the decoder.

### Virtual pipeline data units (VPDUs)

Virtual pipeline data units (VPDUs) are defined as non-overlapping units in a picture. In hardware decoders, successive VPDUs are processed by multiple pipeline stages at the same time. The VPDU size is roughly proportional to the buffer size in most pipeline stages, so it is important to keep the VPDU size small. In most hardware decoders, the VPDU size can be set to maximum transform block (TB) size. However, in VVC, ternary tree (TT) and binary tree (BT) partition may lead to the increasing of VPDUs size.

In order to keep the VPDU size as 64x64 luma samples, the following normative partition restrictions (with syntax signaling modification) are applied in VTM3, as shown in Figure 8:

* TT split is not allowed for a CU with either width or height, or both width and height equal to 128.
* For a 128xN CU with N ≤ 64 (i.e. width equal to 128 and height smaller than 128), horizontal BT is not allowed.
* For an Nx128 CU with N ≤ 64 (i.e. height equal to 128 and width smaller than 128), vertical BT is not allowed.



**Figure 8 – Examples of disallowed TT and BT partitioning in VTM3**

## Intra prediction

### Intra mode coding with 67 intra prediction modes

To capture the arbitrary edge directions presented in natural video, the number of directional intra modes in VTM3 is extended from 33, as used in HEVC, to 65. The new directional modes not in HEVC are depicted as red dotted arrows in Figure 9, and the planar and DC modes remain the same. These denser directional intra prediction modes apply for all block sizes and for both luma and chroma intra predictions.

In VTM3, several conventional angular intra prediction modes are adaptively replaced with wide-angle intra prediction modes for the non-square blocks. Wide angle intra prediction is described in 3.3.1.2.

In HEVC, every intra-coded block has a square shape and the length of each of its side is a power of 2. Thus, no division operations are required to generate an intra-predictor using DC mode. In VTM3, blocks can have a rectangular shape that necessitates the use of a division operation per block in the general case. To avoid division operations for DC prediction, only the longer side is used to compute the average for non-square blocks.

#### Intra mode coding



Figure 9 – 67 intra prediction modes

To keep the complexity of the most probable mode (MPM) list generation low, an intra mode coding method with 6 MPMs is used by considering two available neighboring intra modes. The following three aspects are considered to construct the MPM list:

* + - Default intra modes
    - Neighbouring intra modes
    - Derived intra modes

For neighbor intra modes, two neighbouring blocks, located in left (A) and above (B) are considered.

6 MPM list generation process start with initializing default MPM list as follows:

Default 6 MPM modes = {A, Planar (0) or DC (1), Vertical (50), HOR (18), VER - 4 (46), VER + 4 (54)}

After that 6 MPM modes are updated performing pruning process for two neighboring intra modes. If two neighboring modes are the same each other and the neighboring mode is greater than DC (1) mode, 6 MPM modes are to include three default modes (A, Planar, DC) and three derived modes which are obtained by adding predefined offset values to the neighboring mode and performing modular operation. Otherwise, if two neighboring modes are different, two neighboring modes are assigned to first two MPM modes and the rest four MPM modes are derived from default modes and neighboring modes. During 6 MPM list generation process, pruning is used to remove duplicated modes so that only unique modes can be included into the MPM list. For entropy coding of the 61 non-MPM modes, a Truncated Binary Code (TBC) is used.

#### Wide-angle intra prediction for non-square blocks

Conventional angular intra prediction directions are defined from 45 degrees to -135 degrees in clockwise direction. In VTM3, several conventional angular intra prediction modes are adaptively replaced with wide-angle intra prediction modes for non-square blocks. The replaced modes are signaled using the original mode indexes, which are remapped to the indexes of wide angular modes after parsing. The total number of intra prediction modes is unchanged, i.e., 67, and the intra mode coding method is unchanged.

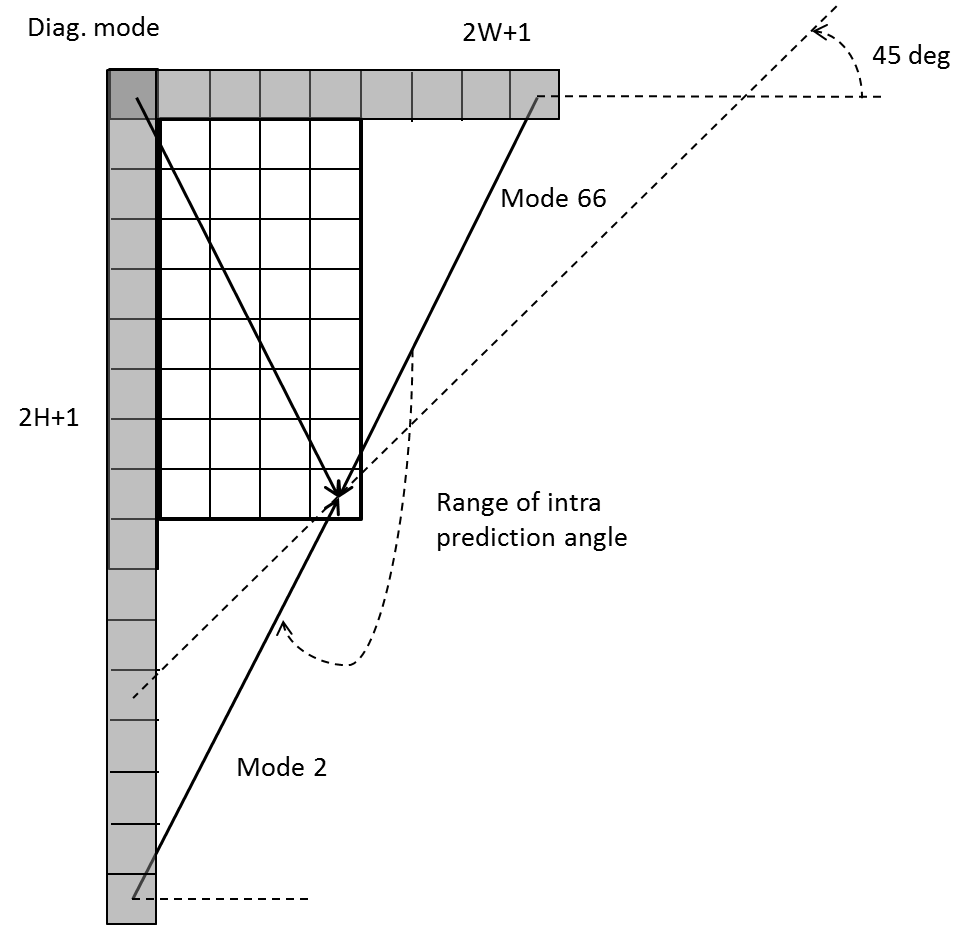
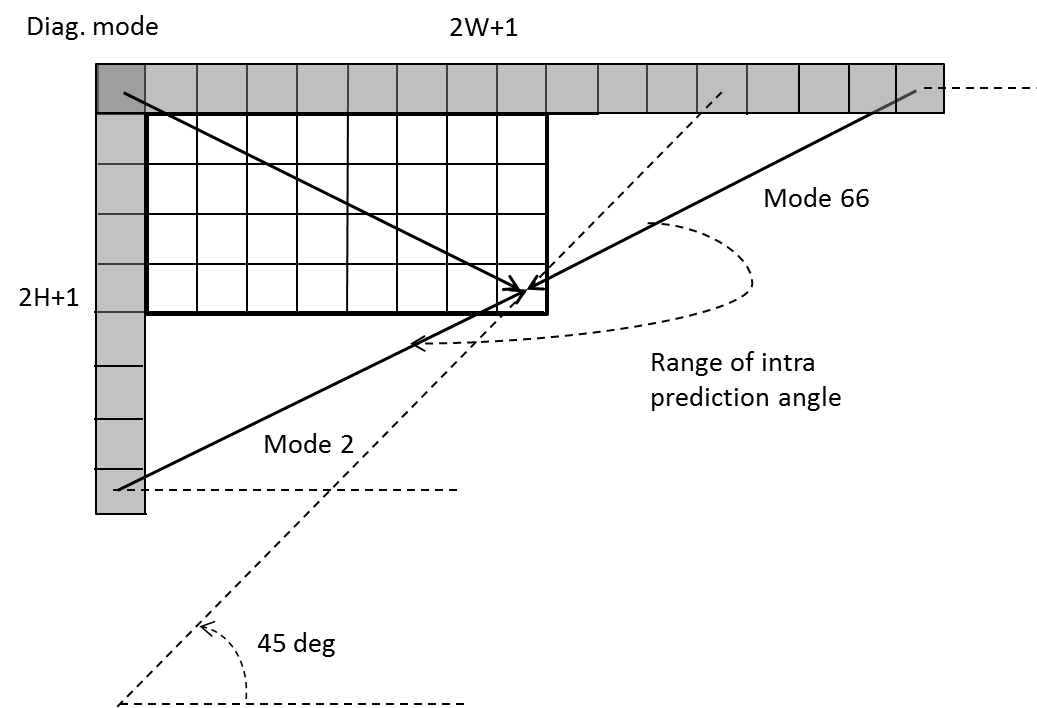


Figure 10 – Reference samples for wide-angular intra prediction

To support these prediction directions, the top reference with length 2W+1, and the left reference with length 2H+1, are defined as shown in Figure 10.

The number of replaced modes in wide-angular direction mode depends on the aspect ratio of a block. The replaced intra prediction modes are illustrated in Table 3‑2

Table 3‑2 - Intra prediction modes replaced by wide-angular modes

|  |  |
| --- | --- |
| Aspect ratio | Replaced intra prediction modes |
| W / H == 16 | Modes 12, 13,14,15 |
| W / H == 8 | Modes 12, 13 |
| W / H == 4 | Modes 2,3,4,5,6,7,8,9,10,11 |
| W / H == 2 | Modes 2,3,4,5,6,7, |
| W / H == 1 | None |
| W / H == 1/2 | Modes 61,62,63,64,65,66 |
| W / H == 1/4 | Mode 57,58,59,60,61,62,63,64,65,66 |
| W / H == 1/8 | Modes 55, 56 |
| W / H == 1/16 | Modes 53, 54, 55, 56 |



Figure 11 - Problem of discontinuity in case of directions beyond 45 degree

As shown in Figure 11, two vertically-adjacent predicted samples may use two non-adjacent reference samples in the case of wide-angle intra prediction. Hence, low-pass reference samples filter and side smoothing are applied to the wide-angle prediction to reduce the negative effect of the increased gap ∆pα.

#### Mode Dependent Intra Smoothing (MDIS)

Four-tap intra interpolation filters are utilized to improve the directional intra prediction accuracy. In HEVC, a two-tap linear interpolation filter has been used to generate the intra prediction block in the directional prediction modes (i.e., excluding Planar and DC predictors). In the VVC 3, simplified 6-bit 4-tap Gaussian interpolation filter is used for only directional intra modes. Non-directional intra prediction process is unmodified. The selection of the 4-tap filters is performed according to the MDIS condition for directional intra prediction modes that provide non-fractional displacements, i.e. to all the directional modes excluding the following: 2, HOR\_IDX, DIA\_IDX, VER\_IDX, 66.

Depending on the intra prediction mode, the following reference samples processing is performed:

1. The directional intra-prediction mode is classified into one of the following groups:
   1. vertical or horizontal modes (HOR\_IDX, VER\_IDX),
   2. diagonal modes that represent angles which are multiple of 45 degree (2, DIA\_IDX, VDIA\_IDX),
   3. remaining directional modes;
2. If the directional intra-prediction mode is classified as belonging to group A, then then no filters are applied to reference samples to generate predicted samples;
3. Otherwise, if a mode falls into group B, then a [1, 2, 1] reference sample filter may be applied (depending on the MDIS condition) to reference samples to further copy these filtered values into an intra predictor according to the selected direction, but no interpolation filters are applied;
4. Otherwise, if a mode is classified as belonging to group C, then only an intra reference sample interpolation filter is applied to reference samples to generate a predicted sample that falls into a fractional or integer position between reference samples according to a selected direction (no reference sample filtering is performed).

#### Cross-component linear model prediction

To reduce the cross-component redundancy, a cross-component linear model (CCLM) prediction mode is used in the VTM3, for which the chroma samples are predicted based on the reconstructed luma samples of the same CU by using a linear model as follows:

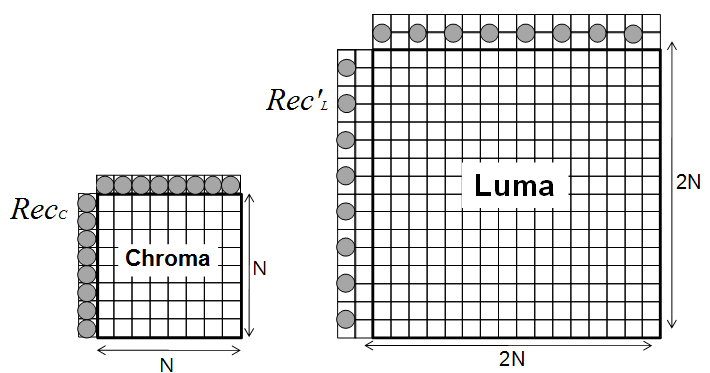
(3-1)

where represents the predicted chroma samples in a CU and represents the downsampled reconstructed luma samples of the same CU. Linear model parameter and are derived from the relation between luma values and chroma values from two samples, which are minimum luma sample A (Xa, Ya) and maximum luma sample B (Xb, Yb) inside the set of neighboring luma samples. The linear model parameters and are obtained according to the following equations.

(3-2)

(3-3)

Where Xa and Ya represent luma value and chroma value of the minimum luma sample. And Xb and Yb  indicate luma value and chroma value of the maximum luma sample, respectively. For a coding block with a square shape, the above two equations are applied directly. For a non-square coding block, the neighbouring samples of the longer boundary are first subsampled to have the same number of samples as for the shorter boundary. Figure 12 shows the location of the left and above samples and the sample of the current block involved in the CCLM mode.



**Figure 12 - Locations of the samples used for the derivation of α and β**

Besides the above template and left template can be used to calculate the linear model coefficients together, they also can be used alternatively in the other 2 LM modes, called LM\_A, and LM\_L modes.

In LM\_A mode, only the above template are used to calculate the linear model coefficients. To get more samples, the above template are extended to (W+H). In LM\_L mode, only left template are used to calculate the linear model coefficients. To get more samples, the left template are extended to (H+W).

For a non-square block, the above template are extended to W+W, the left template are extended to H+H.

Note that only one luma line (general line buffer in intra prediction) is used to make the downsampled luma samples when the upper reference line is at the CTU boundary.

This parameter computation is performed as part of the decoding process, and is not just as an encoder search operation. As a result, no syntax is used to convey the α and β values to the decoder.

For chroma intra mode coding, a total of 8 intra modes are allowed for chroma intra mode coding. Those modes include five traditional intra modes and three cross-component linear model modes (CCLM, LM\_A, and LM\_L). Chroma mode signalling and derivation process are shown in Table 3‑3. Chroma mode coding directly depends on the intra prediction mode of the corresponding luma block. Since separate block partitioning structure for luma and chroma components is enabled in I slices, one chroma block may correspond to multiple luma blocks. Therefore, for Chroma DM mode, the intra prediction mode of the corresponding luma block covering the center position of the current chroma block is directly inherited.

Table 3‑3 – Derivation of chroma prediction mode from luma mode when cclm\_is enabled

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Chroma prediction mode | Corresponding luma intra prediction mode | | | | |
| 0 | 50 | 18 | 1 | X ( 0  <=  X  <=  66 ) |
| 0 | 66 | 0 | 0 | 0 | 0 |
| 1 | 50 | 66 | 50 | 50 | 50 |
| 2 | 18 | 18 | 66 | 18 | 18 |
| 3 | 1 | 1 | 1 | 66 | 1 |
| 4 | 81 | 81 | 81 | 81 | 81 |
| 5 | 82 | 82 | 82 | 82 | 82 |
| 6 | 83 | 83 | 83 | 83 | 83 |
| 7 | 0 | 50 | 18 | 1 | X |

### Position dependent intra prediction combination

In the VTM3, the results of intra prediction of planar mode are further modified by a position dependent intra prediction combination (PDPC) method. PDPC is an intra prediction method which invokes a combination of the un-filtered boundary reference samples and HEVC style intra prediction with filtered boundary reference samples. PDPC is applied to the following intra modes without signalling: planar, DC, horizontal, vertical, bottom-left angular mode and its *eight* adjacent angular modes, and top-right angular mode and its *eight* adjacent angular modes.

The prediction sample *pred*(*x*,*y*) is predicted using an intra prediction mode (DC, planar, angular) and a linear combination of reference samples according to the Equation 3-4 as follows:

*pred*(*x*,*y*)=(*wL*×*R*-1*,y* + *wT*×*Rx,*-1 – *wTL* ×*R*-1*,*-1+(64 – *wL* – *wT*+*wTL*)×*pred*(*x*,*y*) + 32 )>>6 (3-4)

where *Rx,*-1, *R*-1*,y* represent the reference samples located at the top and left of current sample (*x*, *y*), respectively, and *R*-1*,*-1 represents the reference sample located at the top-left corner of the current block.

If PDPC is applied to DC, planar, horizontal, and vertical intra modes, additional boundary filters are not needed, as required in the case of HEVC DC mode boundary filter or horizontal/vertical mode edge filters.

Figure 13 illustrates the definition of reference samples (*Rx,*-1, *R*-1*,y* and *R*-1*,*-1) for PDPC applied over various prediction modes. The prediction sample *pred* (*x’*, *y’*) is located at (*x’*, *y’*) within the prediction block. The coordinate *x* of the reference sample *Rx,*-1 is given by: *x* = *x’* + *y’* + 1, and the coordinate *y* of the reference sample *R*-1*,y* is similarly given by: *y* = *x’* + *y’* + 1.

|  |  |
| --- | --- |
| 1. Diagonal top-right mode | 1. Diagonal bottom-left mode |
| (c) Adjacent diagonal top-right mode | 1. Adjacent diagonal bottom-left mode |

**Figure 13 - Definition of samples used by PDPC applied to diagonal and adjacent angular intra modes.**

The PDPC weights are dependent on prediction modes and are shown in 4.

**Table 3‑4 - Example of PDPC weights according to prediction modes**

|  |  |  |  |
| --- | --- | --- | --- |
| Prediction modes | wT | wL | wTL |
| Diagonal top-right | 16 >> ( ( *y’*<<1 ) >> *shift*) | 16 >> ( ( *x’*<<1 ) >> *shift*) | 0 |
| Diagonal bottom-left | 16 >> ( ( *y’*<<1 ) >> *shift* ) | 16 >> ( ( *x’*<<1 ) >> *shift* ) | 0 |
| Adjacent diagonal top-right | 32 >> ( ( *y’*<<1 ) >> *shift* ) | 0 | 0 |
| Adjacent diagonal bottom-left | 0 | 32 >> ( ( *x’*<<1 ) >> *shift* ) | 0 |

### Multiple reference line (MRL) intra prediction

Multiple reference line (MRL) intra prediction uses more reference lines for intra prediction. In Figure 14, an example of 4 reference lines is depicted, where the samples of segments A and F are not fetched from reconstructed neighbouring samples but padded with the closest samples from Segment B and E, respectively. HEVC intra-picture prediction uses the nearest reference line (i.e., reference line 0). In MRL, 2 additional lines (reference line 1 and reference line 3) are used.

The index of selected reference line (mrl\_idx) is signaled and used to generate intra predictor. For reference line idx, which is greater than 0, only include additional reference line modes in MPM list and only signal mpm index without remaining mode. The reference line index is signaled before intra prediction modes, and Planar and DC modes are excluded from intra prediction modes in case a nonzero reference line index is signaled.



**Figure 14 Example of four reference lines neighboring to a prediction block**

MRL is disabled for the first line of blocks inside a CTU to prevent using extended reference samples outside the current CTU line. Also, PDPC is disabled when additional line is used.

## Inter prediction

For each inter-predicted CU, motion parameters consisting of motion vectors, reference picture indices and reference picture list usage index, and additional information needed for the new coding feature of VVC to be used for inter-predicted sample generation. The motion parameter can be signalled in an explicit or implicit manner. When a CU is coded with skip mode, the CU is associated with one PU and has no significant residual coefficients, no coded motion vector delta or reference picture index. A merge mode is specified whereby the motion parameters for the current CU are obtained from neighbouring CUs, including spatial and temporal candidates, and additional schedules introduced in VVC. The merge mode can be applied to any inter-predicted CU, not only for skip mode. The alternative to merge mode is the explicit transmission of motion parameters, where motion vector, corresponding reference picture index for each reference picture list and reference picture list usage flag and other needed information are signalled explicitly per each CU.

Beyond the inter coding features in HEVC, the VTM3 includes a number of new and refined inter prediction coding tools listed as follows:

* Extended merge prediction
* Merge mode with MVD (MMVD)
* Affine motion compensated prediction
* Subblock-based temporal motion vector prediction (SbTMVP)
* Adaptive motion vector resolution (AMVR)
* Motion field storage: 1/16th luma sample MV storage and 8x8 motion field compression
* Bi-prediction with weighted averaging (BWA)
* Bi-directional optical flow (BDOF)
* Triangle partition prediction
* Combined inter and intra prediction (CIIP)

The following text provides the details on the inter prediction methods specified in VVC.

### Extended merge prediction

In VTM3, the merge candidate list is constructed by including the following five types of candidates in order:

1. Spatial MVP from spatial neighbour CUs
2. Temporal MVP from collocated CUs
3. History-based MVP from an FIFO table
4. Pairwise average MVP
5. Zero MVs.

The size of merge list is signalled in slice header and the maximum allowed size of merge list is 6 in VTM3. For each CU code in merge mode, an index of best merge candidate is encoded using truncated unary binarization (TU). The first bin of the merge index is coded with context and bypass coding is used for other bins.

The generation process of each category of merge candidates is provided in this session.

#### Spatial candidates derivation

The derivation of spatial merge candidates in VVC is same to that in HEVC. A maximum of four merge candidates are selected among candidates located in the positions depicted in **Figure** 15. The order of derivation is A1, B1, B0, A0 and B2. Position B2 is considered onlywhen any CU of position A1, B1, B0, A0 is not available (e.g. because it belongs to another slice or tile) or is intra coded. After candidate at position A1 is added, the addition of the remaining candidates is subject to a redundancy check which ensures that candidates with same motion information are excluded from the list so that coding efficiency is improved. To reduce computational complexity, not all possible candidate pairs are considered in the mentioned redundancy check. Instead only the pairs linked with an arrow in Figure 16 are considered and a candidate is only added to the list if the corresponding candidate used for redundancy check has not the same motion information.



Figure 15– Positions of spatial merge candidate



Figure 16 – Candidate pairs considered for redundancy check of spatial merge candidates

#### Temporal candidates derivation

In this step, only one candidate is added to the list. Particularly, in the derivation of this temporal merge candidate, a scaled motion vector is derived based on co-located CU belonging to the collocatedreferenncee picture. The reference picture list to be used for derivation of the co-located CU is explicitly signalled in the slice header. The scaled motion vector for temporal merge candidate is obtained as illustrated by the dotted line in Figure 17, which is scaled from the motion vector of the co-located CU using the POC distances, tb and td, where tb is defined to be the POC difference between the reference picture of the current picture and the current picture and td is defined to be the POC difference between the reference picture of the co-located picture and the co-located picture. The reference picture index of temporal merge candidate is set equal to zero.

curr\_pic

col\_pic

col\_ref

curr\_ref

td

tb

curr\_CU

col\_CU

Figure 17 – Illustration of motion vector scaling for temporal merge candidate

The position for the temporal candidate is selected between candidates C0 and C1, as depicted in Figure 18. If CU at position C0 is not available, is intra coded, or is outside of the current row of CTUs, position C1 is used. Otherwise, position C0 is used in the derivation of the temporal merge candidate.



Figure 18 – Candidate positions for temporal merge candidate, C0 and C1

#### History-based merge candidates derivation

The history-based MVP (HMVP) merge candidates are added to merge list after the spatial MVP and TMVP. In this method, the motion information of a previously coded block is stored in a table and used as MVP for the current CU. The table with multiple HMVP candidates is maintained during the encoding/decoding process. The table is reset (emptied) when a new CTU row is encountered. Whenever there is a non-subblock inter-coded CU, the associated motion information is added to the last entry of the table as a new HMVP candidate.

In VTM3 the HMVP table size *S* is set to be 6, which indicates up to 6 History-based MVP (HMVP) candidates may be added to the table. When inserting a new motion candidate to the table, a constrained first-in-first-out (FIFO) rule is utilized wherein redundancy check is firstly applied to find whether there is an identical HMVP in the table. If found, the identical HMVP is removed from the table and all the HMVP candidates afterwards are moved forward,

HMVP candidates could be used in the merge candidate list construction process. The latest severalHMVP candidates in the table are checked in order and inserted to the candidate list after the TMVP candidate. Redundancy check is applied on the HMVP candidates to the spatial or temporal merge candidate.

To reduce the number of redundancy check operations, the following simplifications are introduced:

1. Number of HMPV candidates is used for merge list generation is set as (*N* <= 4 ) ? *M*: (8 – *N*), wherein N indicates number of existing candidates in the merge list and M indicates number of available HMVP candidates in the table.
2. Once the total number of available merge candidates reaches the maximally allowed merge candidates minus 1, the merge candidate list construction process from HMVP is terminated.

#### Pair-wise average merge candidates derivation

Pairwise average candidates are generated by averaging predefined pairs of candidates in the existing merge candidate list, and the predefined pairs are defined as {(0, 1), (0, 2), (1, 2), (0, 3), (1, 3), (2, 3)}, where the numbers denote the merge indices to the merge candidate list. The averaged motion vectors are calculated separately for each reference list. If both motion vectors are available in one list, these two motion vectors are averaged even when they point to different reference pictures; if only one motion vector is available, use the one directly; if no motion vector is available, keep this list invalid.

When the merge list is not full after pair-wise average merge candidates are added, the zero MVPs are inserted in the end until the maximum merge candidate number is encountered.

### Merge mode with MVD (MMVD)

In addition to merge mode, where the implicitly derived motion information is directly used for prediction samples generation of the current CU, the merge mode with motion vector differences (MMVD) is introduced in VVC. A MMVD flag is singnaled right after sending a skip flag and merge flag to specify whehther MMVD mode is used for a CU.

In MMVD, after a merge candidate is selected, it is further refined by the signaled MVDs information. The further information includes a merge candidate flag, an index to specify motion magnitude, and an index for indication of motion direction. In MMVD mode, one for the first two candidates in the merge list is selected to be used as MV basis. The merge candidate flag is signaled to specify which one is used.

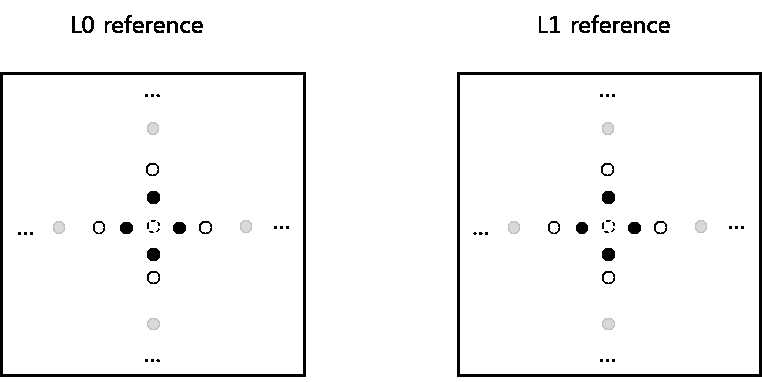


Figure 19 – MMVD Search Point

Distance index specifies motion magnitude information and indicate the pre-defined offset from the starting point. As shown in Figure 19, an offset is added to either horizontal component or vertical component of starting MV. The relation of distance index and pre-defined offset is specified in Table 3‑5

Table 3‑5 – The relation of distance index and pre-defined offset

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Distance IDX** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| **Offset (in unit of luma sample)** | 1/4 | 1/2 | 1 | 2 | 4 | 8 | 16 | 32 |

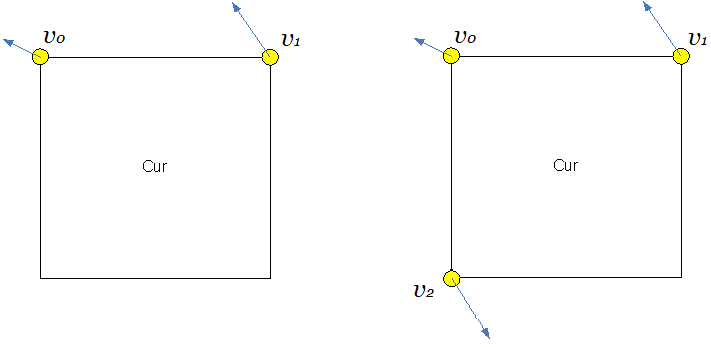
Direction index represents the direction of the MVD relative to the starting point. The direction index can represent of the four directions as shown in Table 3‑6. It’s noted that the meaning of MVD sign could be variant according to the information of starting MVs. When the starting MVs is an un-prediction MV or bi-prediction MVs with both lists point to the same side of the current picture (i.e. POCs of two references are both larger than the POC of the current picture, or are both smaller than the POC of the current picture), the sign in Table 3‑6 specifies the sign of MV offset added to the starting MV. When the starting MVs is bi-prediction MVs with the two MVs point to the different sides of the current picture (i.e. the POC of one reference is larger than the POC of the current picture, and the POC of the other reference is smaller than the POC of the current picture), the sign in Table 3‑6 specifies the sign of MV offset added to the list0 MV component of starting MV and the sign for the list1 MV has opposite value.

Table 3‑6 – Sign of MV offset specified by direction index

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Direction IDX** | 00 | 01 | 10 | 11 |
| **x-axis** | + | – | N/A | N/A |
| **y-axis** | N/A | N/A | + | – |

### Affine motion compensated prediction

In HEVC, only translation motion model is applied for motion compensation prediction (MCP). While in the real world, there are many kinds of motion, e.g. zoom in/out, rotation, perspective motions and the other irregular motions. In the VTM3, a block-based affine transform motion compensation prediction is applied. As shown Figure 20, the affine motion field of the block is described by motion information of two control point (4-parameter) or three control point motion vectors (6-parameter).



1. **4 parameter affine model (b) 6 parameter affine model**

**Figure 20 – control point based affine motion model**

For 4-parameter affine motion model, motion vector at sample location (*x, y*) in a block is derived as:

(3-5)

For 6-parameter affine motion model, motion vector at sample location (*x, y*) in a block is derived as:

(3-6)

Where (*mv0x*, *mv0y*) is motion vector of the top-left corner control point, (*mv1x*, *mv1y*) is motion vector of the top-right corner control point, and (*mv2x*, *mv2y*) is motion vector of the bottom-left corner control point.

In order to simplify the motion compensation prediction, block based affine transform prediction is applied. To derive motion vector of each 4×4 luma sub-block, the motion vector of the center sample of each sub-block, as shown in Figure 21, is calculated according to above equations, and rounded to 1/16 fraction accuracy. Then the motion compensation interpolation filters are applied to generate the prediction of each sub-block with derived motion vector. The sub-block size of chroma-components is also set to be 4×4. The MV of a 4×4 chroma sub-block is calculated as the average of the MVs of the four corresponding 4×4 luma sub-blocks.



Figure 21 – Affine MVF per sub-block

As done for translational motion inter prediction, there are also two affine motion inter prediction modes: affine merge mode and affine AMVP mode.

#### Affine merge prediction

AF\_MERGE mode can be applied for CUs with both width and height larger than or equal to 8. In this mode the CPMVs of the current CU is generated based on the motion information of the spatial neighboring CUs. . There can be up to five CPMVP candidates and an index is signalled to indicate the one to be used for the current CU. The following three types of CPVM candidate are used to form the affine merge candidate list:

1. Inherited affine merge candidates that extrapolated from the CPMVs of the neighbour CUs
2. Constructed affine merge candidates CPMVPs that are derived using the translational MVs of the neighbour CUs
3. Zero MVs

In VTM3, there are maximum two inherited affine candidates, which are derived from affine motion model of the neighboring blocks, one from left neighboring CUs and one from above neighboring CUs. The candidate blocks are shown in Figure 22. For the left predictor, the scan order is A0->A1, and for the above predictor, the scan order is B0->B1->B2. Only the first inherited candidate from each side is selected. No pruning check is performed between two inherited candidates. When a neighboring affine CU is identified, its control point motion vectors are used to derived the CPMVP candidate in the affine merge list of the current CU. As shown in , if the neighbour left bottom block A is coded in affine mode, the motion vectors , and of the top left corner, above right corner and left bottom corner of the CU which contains the block A are attained. When block A is coded with 4-parameter affine model, the two CPMVs of the current CU are calculated according to , and . In case that block A is coded with 6-parameter affine model, the three CPMVs of the current CU are calculated according to , and .



Figure 22 – Locations of inherited affine motion predictors



**Figure 23 – Control point motion vector inheritance**

Constructed affine candidate means the candidate is constructed by combining the neighbor translational motion information of each control point. The motion information for the control points is derived from the specified spatial neighbors and temporal neighbor shown in Figure 24. CPMVk (k=1, 2, 3, 4) represents the k-th control point. For CPMV1, the B2->B3->A2 blocks are checked and the MV of the first available block is used. For CPMV2, the B1->B0 blocks are checked and for CPMV3, the A1->A0 blocks are checked. For TMVP is used as CPMV4 if it’s available.

After MVs of four control points are attained, affine merge candidates are constructed based on those motion information. The following combinations of control point MVs are used to construct in order:

{CPMV1, CPMV2, CPMV3}, {CPMV1, CPMV2, CPMV4}, {CPMV1, CPMV3, CPMV4},  
{CPMV2, CPMV3, CPMV4}, { CPMV1, CPMV2}, { CPMV1, CPMV3}

The combination of 3 CPMVs constructs a 6-parameter affine merge candidate and the combination of 2 CPMVs constructs a 4-parameter affine merge candidate. To avoid motion scaling process, if the reference indices of control points are different, the related combination of control point MVs is discarded.



**Figure 24 –Locations of Candidates position for constructed affine merge mode**

After inherited affine merge candidates and constructed affine merge candidate are checked, if the list is still not full, zero MVs are inserted to the end of the list.

#### Affine AMVP prediction

Affine AMVP mode can be applied for CUs with both width and height larger than or equal to 16. An affine flag in CU level is signalled in the bitstream to indicate whether affine AMVP mode is used and then another flag is signaled to indicate whether 4-parameter affine or 6-parameter affine. In this mode, the difference of the CPMVs of current CU and their predictors CPMVPs is signalled in the bitstream. The affine AVMP candidate list size is 2 and it is generated by using the following four types of CPVM candidate in order:/

1. Inherited affine AMVP candidates that extrapolated from the CPMVs of the neighbour CUs
2. Constructed affine AMVP candidates CPMVPs that are derived using the translational MVs of the neighbour CUs
3. Translational MVs from neighboring CUs
4. Zero MVs

The checking order of inherited affine AMVP candidates is same to the checking order of inherited affine merge candidates. The only difference is that, for AVMP candidate, only the affine CU that has the same reference picture as in current block is considered. No pruning process is applied when inserting an inherited affine motion predictor into the candidate list.

Constructed AMVP candidate is derived from the specified spatial neighbors shown in Figure 24. The same checking order is used as done in affine merge candidate construction. In addition, reference picture index of the neighboring block is also checked. The first block in the checking order that is inter coded and has the same reference picture as in current CUs is used. There is only one When the current CU is coded with 4-parameter affine mode, and and are both availlalbe, they are added as one candidate in the affine AMVP list. When the current CU is coded with 6-parameter affine mode, and all three CPMVs are available, they are added as one candidate in the affine AMVP list. Otherwise, constructed AMVP candidate is set as unavailable.

If affine AMVP list candidates is still less than 2 after inherited affine AMVP candidates and Constructed AMVP candidate are checked, , and will be added, in order, as the translational MVs to predict all control point MVs of the current CU, when available. Finally, zero MVs are used to fill the affine AMVP list if it is still not full.

#### Affine motion information storage

In VTM3, the CPMVs of affine CUs are stored in a separate buffer. The stored CPMVs are only used to generate the inherited CPMVPs in affine merge mode and affine AMVP mode for the lately coded CUs. The sub-block MVs derived from CPMVs are used for motion compensation, MV derivation of merge/AMVP list of translational MVs and de-blocking.

To avoid the picture line buffer for the additional CPMVs, affine motion data inheritance from the CUs from above CTU is treated differently to the inheritance from the normal neighboring CUs. If the candidate CU for affine motion data inheritance is in the above CTU line, the bottom-left and bottom-right sub-block MVs in the line buffer instead of the CPMVs are used for the affine MVP derivation. In this way, the CPMVs are only stored in local buffer. If the candidate CU is 6-parameter affine coded, the affine model is degraded to 4-parameter model. As shown in Figure 25, along the top CTU boundary, the bottom-left and bottom right sub-block motion vectors of a CU are used for affine inheritance of the CUs in bottom CTUs.

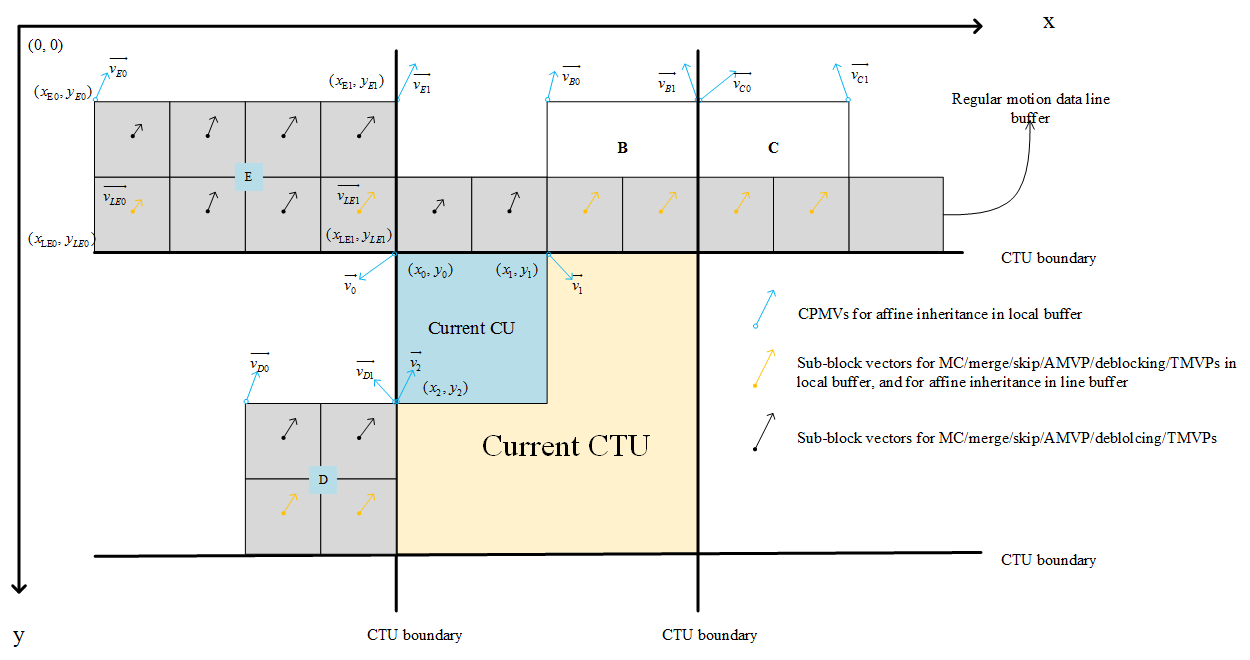


Figure 25 – Illustration of motion vector usage for proposed combined method

### Subblock-based temporal motion vector prediction (SbTMVP)

VTM supports the subblock-based temporal motion vector prediction (SbTMVP) method. Similar to the temporal motion vector prediction (TMVP) in HEVC, SbTMVP uses the motion field in the collocated picture to improve motion vector prediction and merge mode for CUs in the current picture. The same collocated picture used by TMVP is used for SbTVMP. SbTMVP differs from TMVP in the following two main aspects:

1. TMVP predicts motion at CU level but SbTMVP predicts motion at sub-CU level;
2. Whereas TMVP fetches the temporal motion vectors from the collocated block in the collocated picture (the collocated block is the bottom-right or center block relative to the current CU), SbTMVP applies a motion shift before fetching the temporal motion information from the collocated picture, where the motion shift is obtained from the motion vector from one of the spatial neighboring blocks of the current CU.

The SbTVMP process is illustrated in Figure 26. SbTMVP predicts the motion vectors of the sub-CUs within the current CU in two steps. In the first step, the spatial neighbors in Figure 26 (a) are examined in the order of A1, B1, B0 and A0. As soon as and the first spatial neighboring block that has a motion vector that uses the collocated picture as its reference picture is identified, this motion vector is selected to be the motion shift to be applied. If no such motion is identified from the spatial neighbors, then the motion shift is set to (0, 0).

In the second step, the motion shift identified in Step 1 is applied (i.e. added to the current block’s coordinates) to obtain sub-CU-level motion information (motion vectors and reference indices) from the collocated picture as shown in Figure 26 (b). The example in Figure 26 (b) assumes the motion shift is set to block A1’s motion. Then, for each sub-CU, the motion information of its corresponding block (the smallest motion grid that covers the center sample) in the collocated picture is used to derive the motion information for the sub-CU. After the motion information of the collocated sub-CU is identified, it is converted to the motion vectors and reference indices of the current sub-CU in a similar way as the TMVP process of HEVC, where temporal motion scaling is applied to align the reference pictures of the temporal motion vectors to those of the current CU.



1. **Spatial neighboring blocks used by ATVMP**



1. **Deriving sub-CU motion field by applying a motion shift from spatial neighbor and scaling the motion information from the corresponding collocated sub-CUs**

**Figure 26 – The SbTMVP process in VVC**

In VTM3, a combined sub-block based merge list which contains both SbTVMP candidate and affine merge candidates is used for the signalling of sub-block based merge mode. The SbTVMP mode is enabled/disabled by a sequence parameter set (SPS) flag. If the SbTMVP mode is enabled, the SbTMVP predictor is added as the first entry of the list of sub-block based merge candidates, and followed by the affine merge candidates. The size of sub-block based merge list is signalled in SPS and the maximum allowed size of the sub-block based merge list is 5 in VTM3.

The sub-CU size used in SbTMVP is fixed to be 8x8, and as done for affine merge mode, SbTMVP mode is only applicable to the CU with both width and height are larger than or equal to 8.

The encoding logic of the additional SbTMVP merge candidate is the same as for the other merge candidates, that is, for each CU in P or B slice, an additional RD check is performed to decide whether to use the SbTMVP candidate.

### Adaptive motion vector resolution (AMVR)

In HEVC, motion vector differences (MVDs) (between the motion vector and predicted motion vector of a CU) are signalled in units of quarter-luma-sample when use\_integer\_mv\_flag is equal to 0 in the slice header. In VVC, a CU-level adaptive motion vector resolution (AMVR) scheme is introduced. AMVR allows MVD of the CU to be coded in units of quarter-luma-sample, integer-luma-sample or four-luma-sample. The CU-level MVD resolution indication is conditionally signalled if the current CU has at least one non-zero MVD component. If all MVD components (that is, both horizontal and vertical MVDs for reference list L0 and reference list L1) are zero, quarter-luma-sample MVD resolution is inferred.

For a CU that has at least one non-zero MVD component, a first flag is signalled to indicate whether quarter-luma-sample MVD precision is used for the CU. If the first flag is 0, no further signaling is needed and quarter-luma-sample MVD precision is used for the current CU. Otherwise, a second flag is signalled to indicate whether integer-luma-sample or four-luma-sample MVD precision is used. In order to ensure the reconstructed MV has the intended precision (quarter-luma-sample, interger-luma-sample or four-luma-sample), the motion vector predictors for the CU will be rounded to the same precision as that of the MVD before being added together with the MVD. The motion vector predictors are rounded toward zero (that is, a negative motion vector predictor is rounded toward positive infinity and a positive motion vector predictor is rounded toward negative infinity).The encoder determines the motion vector resolution for the current CU using RD check. To avoid always performing CU-level RD check three times for each MVD resolution, in VTM3, the RD check of four-luma-sample MVD resolution is only invoked conditionally. The RD cost of quarter-luma-sample MVD precision is computed first. Then, the RD cost of integer-luma-sample MVD precision is compared to that of quarter-luma-sample MVD precision to decide whether it is necessary to further check the RD cost of four-luma-sample MVD precision. When the RD cost for quarter-luma-sample MVD precision is much smaller than that of the integer-luma-sample MVD precision, the RD check of four-luma-sample MVD precision is skipped.

### Motion field storage

In VTM3, the highest precision of explicitly signalled motion vectors is quarter-luma-sample. In some inter prediction modes such as the affine mode, motion vectors are derived at 1/16th-luma-sample precision and motion compensated prediction is performed at 1/16th-sample-precision. In terms of internal motion field storage, all motion vectors are stored at 1/16th-luma-sample precision.

For temporal motion field storage used by TMVP and ATVMP, motion field compression is performed at 8x8 granularity in contrast to the 16x16 granularity in HEVC.

### Bi-prediction with weighted averaging (BWA)

In HEVC, the bi-prediction signal is generated by averaging two prediction signals obtained from two different reference pictures and/or using two different motion vectors. In VTM-3.0, the bi-prediction mode is extended beyond simple averaging to allow weighted averaging of the two prediction signals.

|  |  |
| --- | --- |
|  | (3-7) |

Five weights are allowed in the weighted averaging bi-prediction, For each bi-predicted CU, the weight w is determined in one of two ways: 1) for a non-merge CU, the weight index is signalled after the motion vector difference; 2) for a merge CU, the weight index is inferred from neighbouring blocks based on the merge candidate index. Weighted averaging bi-prediction is only applied to CUs with 256 or more luma samples (i.e., CU width times CU height is greater than or equal to 256). For low-delay pictures, all 5 weights are used. For non-low-delay pictures, only 3 weights (w∈{3,4,5}) are used.

* 1. At the encoder, fast search algorithms are applied to find the weight index without significantly increasing the encoder complexity. These algorithms are summarized as follows. For further details readers are referred to the VTM software and document JVET-L0646. When combined with AMVR, unequal weights are only conditionally checked for 1-pel and 4-pel motion vector precisions if the current picture is a low-delay picture.
  2. When combined with affine, affine ME will be performed for unequal weights if and only if the affine mode is selected as the current best mode.
  3. When the two reference pictures in bi-prediction are the same, unequal weights are only conditionally checked.
  4. Unequal weights are not searched when certain conditions are met, depending on the POC distance between current picture and its reference pictures, the coding QP, and the temporal level.

### Bi-directional optical flow (BDOF)

The bi-directional optical flow (BDOF) tool is included in VTM-3.0. BDOF, previously referred to as BIO, was included in the JEM. Compared to the JEM version, the BDOF in VTM-3.0 is a simpler version that requires much less computation, especially in terms of number of multiplications and the size of the multiplier.

BDOF is used to refine the bi-prediction signal of a CU at the 4×4 sub-block level. BDOF is applied to a CU if it satisfies the following conditions: 1) the CU’s height is not 4, and the CU is not in size of 4×8, 2) the CU is not coded using affine mode or the ATMVP merge mode; 3) the CU is coded using “true” bi-prediction mode, i.e., one of the two reference pictures is prior to the current picture in display order and the other is after the current picture in display order. BDOF is only applied to the luma component.

As its name indicates, the BDOF mode is based on the optical flow concept, which assumes that the motion of an object is smooth. For each 4×4 sub-block, a motion refinement is calculated by minimizing the difference between the L0 and L1 prediction samples. The motion refinement is then used to adjust the bi-predicted sample values in the 4x4 sub-block. The following steps are applied in the BDOF process.

First, the horizontal and vertical gradients, and , , of the two prediction signals are computed by directly calculating the difference between two neighboring samples, i.e.,

|  |  |
| --- | --- |
|  | (3-8) |

where are the sample value at coordinate of the prediction signal in list , .

Then, the auto- and cross-correlation of the gradients, , , , and , are calculated as

|  |  |
| --- | --- |
| , | (3-9) |

where

|  |  |
| --- | --- |
|  | (3-10) |

where is a 6×6 window around the 4×4 sub-block.

The motion refinement is then derived using the cross- and auto-correlation terms using the following:

|  |  |
| --- | --- |
|  | (3-11) |

where , , . and is the floor function.

Based on the motion refinement and the gradients, the following adjustment is calculated for each sample in the 4×4 sub-block:

|  |  |
| --- | --- |
|  | (3-12) |

Finally, the BDOF samples of the CU are calculated by adjusting the bi-prediction samples as follows:

|  |  |
| --- | --- |
|  | (3-13) |

In the above, the values of , and are equal to 3, 6, and 12, respectively. These values are selected such that the multipliers in the BDOF process do not exceed 15-bit, and the maximum bit-width of the intermediate parameters in the BDOF process is kept within 32-bit.

In order to derive the gradient values, some prediction samples in list () outside of the current CU boundaries need to be generated. As depicted in **Figure 27**, the BDOF in VTM-3.0 uses one extended row/column around the CU’s boundaries. In order to control the computational complexity of generating the out-of-boundary prediction samples, bilinear filter is used to generate prediction samples in the extended area (white positions), and the normal 8-tap motion compensation interpolation filter is used to generate prediction samples within the CU (gray positions). These extended sample values are used in gradient calculation only. For the remaining steps in the BDOF process, if any sample and gradient values outside of the CU boundaries are needed, they are padded (i.e. repeated) from their nearest neighbors.



**Figure 27 – Extended CU region used in BDOF**

### Triangle partition for inter prediction

In VTM3, a new triangle partition mode is introduced for inter prediction. The triangle partition mode is only applied to CUs that are 8x8 or larger and are coded in skip or merge mode. For a CU satisfying these conditions, a CU-level flag is signalled to indicate whether the triangle partition mode is applied or not.

When this mode is used, a CU is split evenly into two triangle-shaped partitions, using either the diagonal split or the anti-diagonal split (Figure 28). Each triangle partition in the CU is inter-predicted using its own motion; only uni-prediction is allowed for each partition, that is, each partition has one motion vector and one reference index. The uni-prediction motion constraint is applied to ensure that same as the conventional bi-prediction, only two motion compensated prediction are needed for each CU. The uni-prediction motion for each partition is derived from a uni-prediction candidate list constructed using the process in 3.4.9.1.



**Figure 28 – Triangle partition based inter prediction**

If the CU-level flag indicates that the current CU is coded using the triangle partition mode, an index in the range of [0, 39] is further signalled. Using this triangle partition index, the direction of the triangle partition (diagonal or anti-diagonal), as well as the motion for each of the partitions can be obtained through a look-up table. After predicting each of the triangle partitions, the sample values along the diagonal or anti-diagonal edge are adjusted using a blending processing with adaptive weights. This is the prediction signal for the whole CU, and transform and quantization process will be applied to the whole CU as in other prediction modes. Finally, the motion field of a CU predicted using the triangle partition mode is stored in 4x4 units as in 3.4.9.3.

#### Uni-prediction candidate list construction

The uni-prediction candidate list consists of five uni-prediction motion vector candidates. It is derived from seven neighboring blocks including five spatial neighboring blocks (labelled 1 to 5 in Figure 29) and two temporal co-located blocks (labelled 6 to 7 in Figure 29). The motion vectors of the seven neighboring blocks are collected and put into the uni-prediction candidate list according to the following order: first, the motion vectors of the uni-predicted neighboring blocks; then, for the bi-predicted neighboring blocks, the L0 motion vectors (that is, the L0 motion vector part of the bi-prediction MV), the L1 motion vectors (that is, the L1 motion vector part of the bi-prediction MV), and averaged motion vectors of the L0 and L1 motion vectors of the bi-prediction MVs. If the number of candidates is less than five, zero motion vector is added to the end of the list.



**Figure 29 – Spatial and temporal neighboring blocks used to construct the uni-prediction candidate list**

There are 40 possible ways to prediction a CU coded in triangle partition mode: 5 (for partition 1 motion) x 4 (for partition 2 motion) x 2 (diagonal or anti-diagonal partition modes). The triangle partition index in the range of [0, 39] is used to identify which one of these possibilities is used using the look-up table in Table 3‑7.

Table 3‑7 – Look up table used to derive triangle direction and partition motions based on triangle index

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| triangle\_idx | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| triangle dir | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| Part 1 cand | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 1 | 3 | 4 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 2 |
| Part 2 cand | 0 | 1 | 2 | 1 | 0 | 3 | 4 | 0 | 0 | 0 | 2 | 2 | 2 | 4 | 3 | 3 | 4 | 4 | 3 | 1 |
| triangle\_idx | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |
| triangle dir | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Part 1 cand | 2 | 2 | 4 | 3 | 3 | 3 | 4 | 3 | 2 | 4 | 4 | 2 | 4 | 3 | 4 | 3 | 2 | 2 | 4 | 3 |
| Part 2 cand | 0 | 1 | 3 | 0 | 2 | 4 | 0 | 1 | 3 | 1 | 1 | 3 | 2 | 2 | 3 | 1 | 4 | 4 | 2 | 4 |

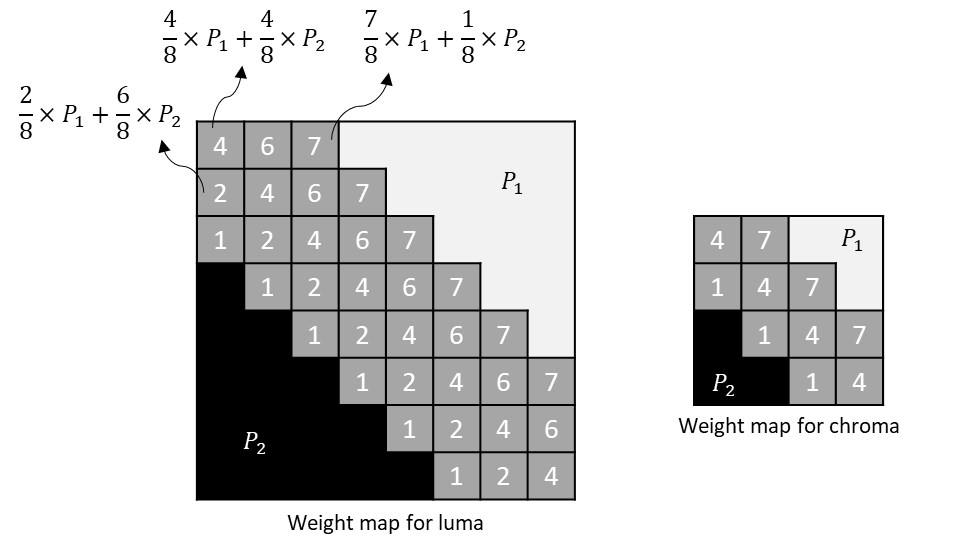
#### Blending along the triangle partition edge

After predicting each triangle partition using its own motion, blending is applied to the two prediction signals to derive samples around the diagonal or anti-diagonal edge. The blending process adaptively chooses between two sets of weights depending on the motion vector difference between the two partitions. The two weight sets are as follows:

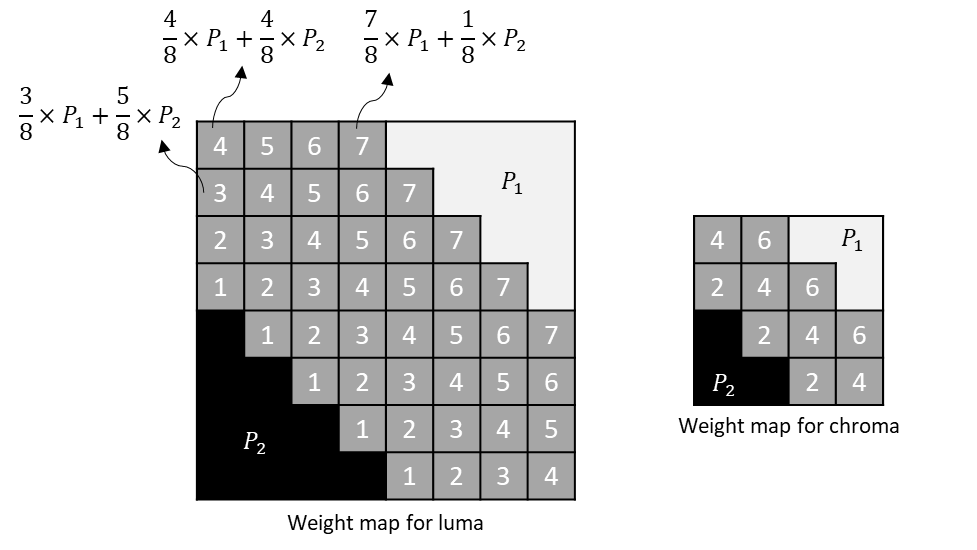
* 1st set: {7/8, 6/8, 4/8, 2/8, 1/8} for luma and {7/8, 4/8, 1/8} for chroma, Figure 30 (a);
* 2nd set: {7/8, 6/8, 5/8, 4/8, 3/8, 2/8, 1/8} for luma and {6/8, 4/8, 2/8} for chroma, Figure 30 (b).

The second set has more luma weights and blends more luma samples along the partition edge. The following condition is used to select the weight set:

* If the reference pictures of the two triangle partitions are different from each other, or if their motion vector difference is larger than 16 luma samples, then select the 2nd set;
* Otherwise, select the 1st set.



* + 1. First weight set



* + 1. Second weight set

**Figure 30 – weight sets used in the blending process**

#### Motion field storage

The motion vectors of a CU coded in triangle partition mode are stored in 4x4 units. Depending on the position of each 4x4 unit, either uni-prediction or bi-prediction motion vectors are stored. Denote Mv1 and Mv2 as uni-prediction motion vectors for partition 1 and partition 2, respectively. If a 4x4 unit is located in the non-weighted area shown in the example of Figure 30, either Mv1 or Mv2 is stored for that 4x4 unit. Otherwise, if the 4x4 unit is located in the weighted area, a bi-prediction motion vector is stored. The bi-prediction motion vector is derived from Mv1 and Mv2 according to the following process:

1. If Mv1 and Mv2 are from different reference picture lists (one from L0 and the other from L1), then Mv1 and Mv2 are simply combined to form the bi-prediction motion vector.
2. Otherwise, if Mv1 and Mv2 are from the same list, and without loss of generality, assume they are both from L0. In this case,
   1. If the reference picture of either Mv2 (or Mv1) appears in L1, then that Mv2 (or Mv1) is converted to a L1 motion vector using that reference picture in L1. Then the two motion vectors are combined to form the bi-prediction motion vector;
   2. Otherwise, instead of bi-prediction motion, only uni-prediction motion Mv1 is stored.

### Combined inter and intra prediction (CIIP)

In VTM3, when a CU is coded in merge mode, and if the CU contains at least 64 luma samples (that is, CU width times CU height is equal to or larger than 64), an additional flag is signalled to indicate if the combined inter/intra prediction (CIIP) mode is applied to the current CU.

In order to form the CIIP prediction, an intra prediction mode is first derived from two additional syntax elements. Up to four possible intra prediction modes can be used: DC, planar, horizontal, or vertical. Then, the inter prediction and intra prediction signals are derived using regular intra and inter decoding processes. Finally, weighted averaging of the inter and intra prediction signals is performed to obtain the CIIP prediction.

#### Intra prediction mode derivation

Up to 4 intra prediction modes, including DC, PLANAR, HORIZONTAL, and VERTICAL modes, can be used to predict the luma component in the CIIP mode. If the CU shape is very wide (that is, width is more than two times of height), then the HORIZONTAL mode is not allowed. If the CU shape is very narrow (that is, height is more than two times of width), then the VERTICAL mode is not allowed. In these cases, only 3 intra prediction modes are allowed.

The CIIP mode uses 3 most probable modes (MPM) for intra prediction. The CIIP MPM candidate list is formed as follows:

* The left and top neighbouring blocks are set as A and B, respectively
* The intra prediction modes of block A and block B, denoted as intraModeA and intraModeB, respectively, are derived as follows:
  + Let X be either A or B
  + intraModeX is set to DC if 1) block X is not available; or 2) block X is not predicted using the CIIP mode or the intra mode; 3) block B is outside of the current CTU
  + otherwise, intraModeX is set to 1) DC or PLANAR if the intra prediction mode of block X is DC or PLANAR; or 2) VERTICAL if the intra prediction mode of block X is a “vertical-like” angular mode (larger than 34), or 3) HORIZONTAL if the intra prediction mode of block X is a “horizontal-like” angular mode (smaller than or equal to 34)
* If intraModeA and intraModeB are the same:
  + If intraModeA is PLANAR or DC, then the three MPMs are set to {PLANAR, DC, VERTICAL} in that order
  + Otherwise, the three MPMs are set to {intraModeA, PLANAR, DC} in that order
* Otherwise (intraModeA and intraModeB are different):
  + The first two MPMs are set to {intraModeA, intraModeB} in that order
  + Uniqueness of PLANAR, DC and VERTICAL is checked in that order against the first two MPM candidate modes; as soon as a unique mode is found, it is added to as the third MPM

If the CU shape is very wide or very narrow as defined above, the MPM flag is inferred to be 1 without signalling. Otherwise, an MPM flag is signalled to indicate if the CIIP intra prediction mode is one of the CIIP MPM candidate modes.

If the MPM flag is 1, an MPM index is further signalled to indicate which one of the MPM candidate modes is used in CIIP intra prediction. Otherwise, if the MPM flag is 0, the intra prediction mode is set to the “missing” mode in the MPM candidate list. For example, if the PLANAR mode is not in the MPM candidate list, then PLANAR is the missing mode, and the intra prediction mode is set to PLANAR. Since 4 possible intra prediction modes are allowed in CIIP, and the MPM candidate list contains only 3 intra prediction modes, one of the 4 possible modes must be the missing mode.

For the chroma components, the DM mode is always applied without additional signalling; that is, chroma uses the same prediction mode as luma.

The intra prediction mode of a CIIP-coded CU will be saved and used in the intra mode coding of the future neighbouring CUs.

#### Combining the inter and intra prediction signals

The inter prediction signal in the CIIP mode is derived using the same inter prediction process applied to regular merge mode; and the intra prediction signal is derived using the CIIP intra prediction mode following the regular intra prediction process. Then, the intra and inter prediction signals are combined using weighted averaging, where the weight value depends on the intra prediction mode and where the sample is located in the coding block, as follows:

* If the intra prediction mode is the DC or planar mode, or if the block width or height is smaller than 4, then equal weights are applied to the intra prediction and the inter prediction signals.
* Otherwise, the weights are determined based on the intra prediction mode (either horizontal mode or vertical mode in this case) and the sample location in the block. Take the horizontal prediction mode for example (the weights for the vertical mode are derived similarly but in the orthogonal direction). Denote W as the width of the block and H as the height of the block. The coding block is first split into four equal-area parts, each of the dimension (W/4)xH. Starting from the part closest to the intra prediction reference samples and ending at the part farthest away from the intra prediction reference samples, the weight wt for each of the 4 regions is set to 6, 5, 3, and 2, respectively. The final CIIP prediction signal is derived using the following:

|  |  |
| --- | --- |
|  | (3-14) |

### Miscellaneous inter prediction aspects

To reduce memory bandwidth, bi-prediction is not allowed for 4x4 CUs in VVC.

## Transform and quantization

### Large block-size transforms with high-frequency zeroing

In VTM3, large block-size transforms, up to 64×64 in size, are enabled, which is primarily useful for higher resolution video, e.g., 1080p and 4K sequences. High frequency transform coefficients are zeroed out for the transform blocks with size (width or height, or both width and height) equal to 64, so that only the lower-frequency coefficients are retained. For example, for an M×N transform block, with M as the block width and N as the block height, when M is equal to 64, only the left 32 columns of transform coefficients are kept. Similarly, when N is equal to 64, only the top 32 rows of transform coefficients are kept. When transform skip mode is used for a large block, the entire block is used without zeroing out any values.

### Multiple transform selection (MTS) for core transform

In addition to DCT-II which has been employed in HEVC, a Multiple Transform Selection (MTS) scheme is used for residual coding both inter and intra coded blocks. It uses multiple selected transforms from the DCT8/DST7. The newly introduced transform matrices are DST-VII and DCT-VIII. Table 3‑8Table 3‑8 shows the basis functions of the selected DST/DCT.

Table 3‑8 - Transform basis functions of DCT-II/ VIII and DSTVII for N-point input

|  |  |
| --- | --- |
| Transform Type | Basis function *Ti*(*j*), *i*, *j* = 0, 1,…, *N*−1 |
| DCT-II | where, |
| DCT-VIII |  |
| DST-VII |  |

In order to keep the orthogonality of the transform matrix, the transform matrices are quantized more accurately than the transform matrices in HEVC. To keep the intermediate values of the transformed coefficients within the 16-bit range, after horizontal and after vertical transform, all the coefficients are to have 10-bit.

In order to control MTS scheme, separate enabling flags are specified at SPS level for intra and inter, respectively. When MTS is enabled at SPS, a CU level flag is signalled to indicate whether MTS is applied or not. Here, MTS is applied only for luma. The MTS CU level flag is signalled when the following conditions are satisfied.

* + - Both width and height smaller than or equal to 32
    - CBF flag is equal to one

If MTS CU flag is equal to zero, then DCT2 is applied in both directions. However, if MTS CU flag is equal to one, then two other flags are additionally signalled to indicate the transform type for the horizontal and vertical directions, respectively. Transform and signalling mapping table as shown in Table 3‑9. When it comes to transform matrix precision, 8-bit primary transform cores are used. Therefore, all the transform cores used in HEVC are kept as the same, including 4-point DCT-2 and DST-7, 8-point, 16-point and 32-point DCT-2. Also, other transform cores including 64-point DCT-2, 4-point DCT-8, 8-point, 16-point, 32-point DST-7 and DCT-8, use 8-bit primary transform cores.

Table 3‑9 - Transform and signalling mapping table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MTS\_CU\_flag | MTS\_Hor\_flag | MTS\_Ver\_flag | Intra/inter | |
|  |  |  | Horizontal | Vertical |
| 0 |  |  | DCT2 | |
| 1 | 0 | 0 | DST7 | DST7 |
| 0 | 1 | DCT8 | DST7 |
| 1 | 0 | DST7 | DCT8 |
| 1 | 1 | DCT8 | DCT8 |

As in HEVC, the residual of a block can be coded with transform skip mode. To avoid the redundancy of syntax coding, the transform skip flag is not signalled when the CU level MTS\_CU\_flag is not equal to zero. Transform skip is enabled when both block width and height are equal to or less than 4.

### Quantization

In VTM 3, Maximum QP was extended from 51 to 63, and the signaling of initial QP was changed accordingly. The initial value of SliceQpY is modified at the slice segment layer when a non-zero value of slice\_qp\_delta is coded. Specifically, the value of init\_qp\_minus26 is modified to be in the range of −( 26 + QpBdOffsetY ) to +37.

In addition, the same HEVC scalar quantization is used with a new concept called dependent scala quantization. Dependent scalar quantization refers to an approach in which the set of admissible reconstruction values for a transform coefficient depends on the values of the transform coefficient levels that precede the current transform coefficient level in reconstruction order. The main effect of this approach is that, in comparison to conventional independent scalar quantization as used in HEVC, the admissible reconstruction vectors are packed denser in the N-dimensional vector space (N represents the number of transform coefficients in a transform block). That means, for a given average number of admissible reconstruction vectors per N-dimensional unit volume, the average distortion between an input vector and the closest reconstruction vector is reduced. The approach of dependent scalar quantization is realized by: (a) defining two scalar quantizers with different reconstruction levels and (b) defining a process for switching between the two scalar quantizers.



Figure 31 – Illustration of the two scalar quantizers used in the proposed approach of dependent quantization.

The two scalar quantizers used, denoted by Q0 and Q1, are illustrated in **Error! Reference source not found.**. The location of the available reconstruction levels is uniquely specified by a quantization step size Δ. The scalar quantizer used (Q0 or Q1) is not explicitly signalled in the bitstream. Instead, the quantizer used for a current transform coefficient is determined by the parities of the transform coefficient levels that precede the current transform coefficient in coding/reconstruction order.



Figure 32 – State transition and quantizer selection for the proposed dependent quantization.

As illustrated in **Error! Reference source not found.**, the switching between the two scalar quantizers (Q0 and Q1) is realized via a state machine with four states. The state can take four different values: 0, 1, 2, 3. It is uniquely determined by the parities of the transform coefficient levels preceding the current transform coefficient in coding/reconstruction order. At the start of the inverse quantization for a transform block, the state is set equal to 0. The transform coefficients are reconstructed in scanning order (i.e., in the same order they are entropy decoded). After a current transform coefficient is reconstructed, the state is updated as shown in Figure 18, where k denotes the value of the transform coefficient level.

## Entropy coding

In the VVC 3, CABAC contains the following major changes compared to the design in HEVC:

* Transform coefficient coding with five passes in a subblock
* Context modeling for transform coefficients

### Transform coefficient level coding

In HEVC, transform coefficients of a coding block are coded using non-overlapped coefficient groups (or subblocks), and each CG contains the coefficients of a 4x4 block of a coding block. The CGs inside a coding block, and the transform coefficients within a CG, are coded according to pre-defined scan orders. The coding of transform coefficient levels of a CG with at least one non-zero transform coefficient may be separated into multiple scan passes. In the first pass, the first bin (denoted by bin0, also referred as *significant\_coeff\_flag,* which indicates the magnitude of the coefficient is larger than 0) is coded. Next, two scan passes for context coding the second/third bins (denoted by bin1 and bin2, respectively, also referred as *coeff\_abs\_greater1\_flag* and *coeff\_abs\_greater2\_flag)* may be applied. Finally, two more scan passes for coding the sign information and the remaining values (also referred as *coeff\_abs\_level\_remaining)* of coefficient levels are invoked, if necessary. Note that only bins in the first three scan passes are coded in a regular mode and those bins are termed regular bins in the following descriptions.

In the VVC 3, for each subblock, the regular coded bins and the bypass coded bins are separated in coding order; first all regular coded bins for a subblock are transmitted and, thereafter, the bypass coded bins are transmitted. The transform coefficient levels of a subblock are coded in five passes over the scan positions as follows:

* + **Pass 1**: coding of significance (sig\_flag), greater 1 flag (gt1\_flag), and parity (par\_level\_flag) is processed in coding order. If sig\_flag is equal to 1, first the gt1\_flag is coded (which specifies whether the absolute level is greater than 1). If gt1\_flag is equal to 1, the par\_flag is additionally coded (it specifies the parity of the absolute level minus 2).
  + **Pass 2**: coding of greater 2 flags (gt2\_flag) is processed for all scan positions with gt1\_flag equal to 1.
  + **Pass 3**: coding of remaining absolute level (remainder) is processed for all scan positions with gt2\_flag equal to 1 or gt1\_flag equal to 1. The non-binary syntax element is binarized with Golomb-Rice code and the resulting bins are coded in the bypass mode of the arithmetic coding engine.
  + **Pass 4**: absolute level (absLevel) of the coefficients for which no sig\_flag is coded in the first pass (due to reaching the limit of regular-coded bins) are completely coded in the bypass mode of the arithmetic coding engine using a Golomb-Rice code.
  + **Pass 5**: coding of the signs (sign\_flag) for all scan positions with sig\_coeff\_flag equal to 1

It is guaranteed that no more than 32 regular-coded bins (28 bins in pass 1 and 4 bins in pass 2) have to be encoded or decoded for a subblock. For 2x2 chroma subblocks, the number of gt2\_flag’s per subblock is limited to a maximum value of 2 and the number of bins in the first pass for a subblock (sig\_flag, par\_flag, and gt1\_flag) is limited to a maximum number of 6.

The Rice parameter (ricePar) for coding the non-binary syntax element remainder (in Pass 3) is derived similar to HEVC. At the start of each subblock, ricePar is set equal to 0. After coding a syntax element remainder, the Rice parameter is modified according to predefined equation. For coding the non-binary syntax element absLevel (in Pass 4), the sum of absolute values sumAbs in a local template is determined. The variables ricePar and posZero are determined based on dependent quantization and sumAbs by a table look-up. The intermediate variable codeValue is derived as follows:

* + If absLevel[k] is equal to 0, codeValue is set equal to posZero;
  + Otherwise, if absLevel[k] is less than or equal to posZero, codeValue is set equal to absLevel[k] – 1;
  + Otherwise (absLevel[k] is greater than posZero), codeValue is set equal to absLevel[k].

The value of codeValue is coded using a Golomb-Rice code with Rice parameter ricePar.

### Context modeling for coefficient coding

The selection of probability models for the syntax elements related to absolute values of transform coefficient levels depends on the values of the absolute levels or partially reconstructed absolute levels in a local neighbourhood. The template used is illustrated in Figure 33.



Figure 33: Illustration of the template used for selecting probability models. The black square specifies the current scan position and the blue squares represent the local neighbourhood used.

The selected probability models depend on the sum of the absolute levels (or partially reconstructed absolute levels) in a local neighbourhood and the number of absolute levels greater than 0 (given by the number of sig\_coeff\_flags equal to 1) in the local neighbourhood. The context modelling and binarization depends on the following measures for the local neighbourhood:

* numSig: the number of non-zero levels in the local neighbourhood;
* sumAbs1: the sum of partially reconstructed absolute levels (absLevel1) after the first pass   
   in the local neighbourhood;
* sumAbs: the sum of reconstructed absolute levels in the local neighbourhood
* diagonal position (d): the sum of the horizontal and vertical coordinates of a current scan position inside the transform block

Based on the values of numSig, sumAbs1, and d, the probability models for coding sig \_flag, par \_flag, gt1\_flag, and gt2\_flag are selected. The Rice parameter for binarizing abs\_remainder is selected based on the values of sumAbs and numSig.

## In-loop filter

There are totally three in loop filters in VTM3. Besides deblocking filter and SAO (the two loop filters in HEVC), adaptive loop fitler (ALF) are applied in the VTM3. The order of the filtering process in the VTM3 is the deblocking filter, SAO and ALF.

In the VTM3, the SAO and deblocking filtering processes are almost same to those in HEVC.

### In-loop filter

In the VTM3, an adaptive loop filter (ALF) with block-based filter adaption is applied. For the luma component, one among 25 filters is selected for each 4×4 block, based on the direction and activity of local gradients.

#### Filter shape

In the JEM, two diamond filter shapes (as shown in Figure 34) are used. for the luma component. The 7×7 diamond shape is applied for luma component and the 5×5 diamond shape applied for chroma component.

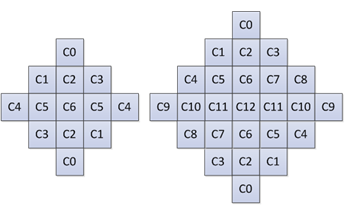


Figure 34 – ALF filter shapes (chroma: 5×5 diamond, luma: 7×7 diamond)

#### Block classification

For luma component, each block is categorized into one out of 25 classes. The classification index *C* is derived based on its directionality and a quantized value of activity , as follows:

(3-15)

To calculate and , gradients of the horizontal, vertical and two diagonal direction are first calculated using 1-D Laplacian:

(3-16)

(3-17)

(3-18)

(3-19)

Where indices and refer to the coordinates of the upper left sample within the block and indicates a reconstructed sample at coordinate .

To reduce the complexity of block classification, the subsampled 1-D Laplacian calculation is applied. As shown in Figure 35, the same subsampled positions are used for gradient calculation of all directions.

|  |  |
| --- | --- |
|  |  |
| (a) Subsampled positions for vertical gradient | (b) Subsampled positions for horizontal gradient |
|  |  |
| (c) Subsampled positions for diagonal gradient | (d) Subsampled positions for diagonal gradient |

**Figure 35 – Subsampled Laplacian calculation**

Then maximum and minimum values of the gradients of horizontal and vertical directions are set as:

, (3-20)

The maximum and minimum values of the gradient of two diagonal directions are set as:

, (3-21)

To derive the value of the directionality , these values are compared against each other and with two thresholds and :

**Step 1**. If both and are true, is set to .

**Step 2**. If , continue from Step 3; otherwise continue from Step 4.

**Step 3**. If , is set to ; otherwise is set to .

**Step 4**. If , is set to ; otherwise is set to .

The activity value is calculated as:

(3-22)

is further quantized to the range of 0 to 4, inclusively, and the quantized value is denoted as .

For chroma components in a picture, no classification method is applied, i.e. a single set of ALF coefficients is applied for each chroma component.

#### Geometric transformations of filter coefficients

Before filtering each 4×4 luma block, geometric transformations such as rotation or diagonal and vertical flipping are applied to the filter coefficients depending on gradient values calculated for that block. This is equivalent to applying these transformations to the samples in the filter support region. The idea is to make different blocks to which ALF is applied more similar by aligning their directionality.

Three geometric transformations, including diagonal, vertical flip and rotation are introduced:

Diagonal: (3-23)

Vertical flip: (3-24)

Rotation: (3-25)

where is the size of the filter and are coefficients coordinates, such that location is at the upper left corner and location is at the lower right corner. The transformations are applied to the filter coefficients *f* (*k*, *l*) depending on gradient values calculated for that block. The relationship between the transformation and the four gradients of the four directions are summarized in the following table.

**Table 3‑10 - Mapping of the gradient calculated for one block and the transformations**

|  |  |
| --- | --- |
| Gradient values | Transformation |
| gd2 < gd1 and gh < gv | No transformation |
| gd2 < gd1 and gv < gh | Diagonal |
| gd1 < gd2 and gh < gv | Vertical flip |
| gd1 < gd2 and gv < gh | Rotation |

#### Filter parameters signalling

In the VTM3, ALF filter parameters are signalled in the slice header. Up to 25 sets of luma filter coefficients could be signalled. To reduce bits overhead, filter coefficients of different classification can be merged.

The filtering process can be controlled at CTB level. A flag is always signalled to indicate whether ALF is applied to a luma CTB. For each chroma CTB, a flag might be signalled to indicate whether ALF is applied to a chroma CTB depends on the value of alf\_chroma\_ctb\_present\_flag.

The filter coefficients are quantized with norm equal to 128. To further restrict the multiplication complexity, a bitstream conformance is applied that the coefficient value of the central position shall be in the range of 0 to 28 and he coefficient values of the remaining positions shall be in the range of −27 to 27 − 1, inclusive.

#### Filtering process

At decoder side, when ALF is enabled for a CTB, each sample within the CU is filtered, resulting in sample value as shown below, where *L* denotes filter length, represents filter coefficient, and denotes the decoded filter coefficients.

(3-26)

### Deblocking filter

In the VTM3, deblocking filtering process is mostly the same to those in HEVC. However, the following modifications are added.

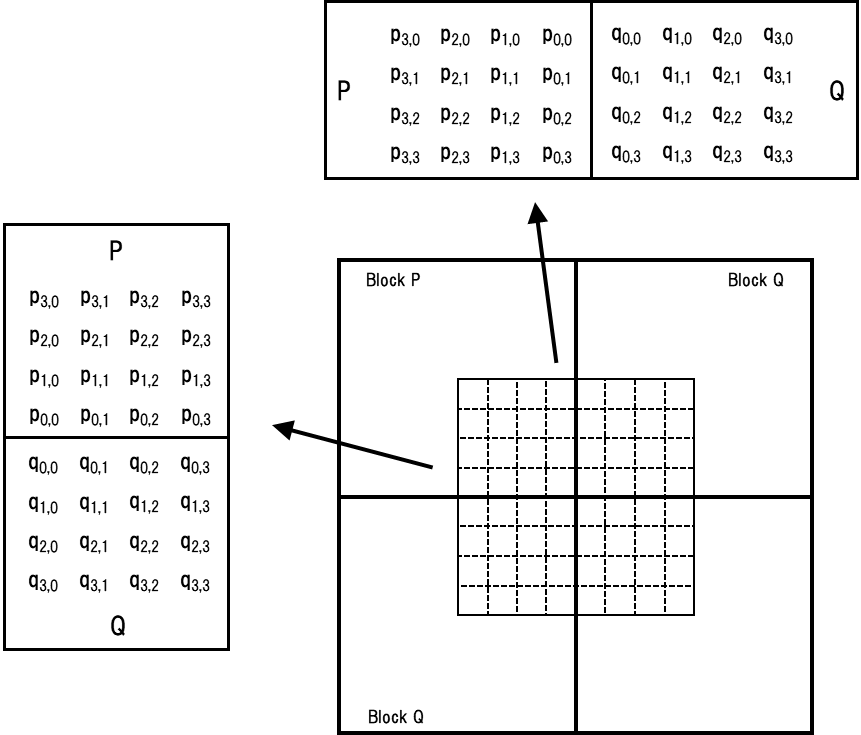
* 1. The filter strength of the deblocking filter dependent of the averaged luma level of the reconstructed samples.
  2. Deblocking tC table extension

#### Filter strength dependent on reconstructed average luma level

In HEVC, the filter strength of the deblocking filter is controlled by the variables β and tC which are derived from the averaged quantization parameters qPL. In the VTM3, deblocking filter controls the strength of the deblocking filter by adding offset to qPL according to the luma level of the reconstructed samples. The reconstructed luma level LL is derived as follow:

LL= ( ( p0,0 + p0,3 + q0,0 + q0,3 ) >> 2 ) / ( 1 << iBitDepth ),

where, the sample values pi,k and qi,k with i = 0..3 and k = 0 and 3 are derived as shown in Figure 28.



**Figure 28 Sample position of pi,k and qi,k**

The variable qPL is derived as follows:

qPL = ( ( QpQ + QpP +1 ) >> 1 ) + qpOffset,

where QpQ and QpP denote the quantization parameters of the coding units containing the sample q0,0 and p0,0, respectively. The offset qpOffset dependent on transfer function, the values are signaled in the SPS.

#### Deblocking tC table extension

In VTM 3, Maximum QP was changed from 51 to 63, and it is desired to reflect corresponding change to deblocking table, which derive values of deblocking parameters tC based on the block QP, The following is updated tC table to accommodate the extension of the QP range.

tC = [ 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6, 7, 8, 9,10,11,13,14,16,18,20,22,25,28,31,35,39,44,50,56,63,70,79,88,99]

#### Deblocking filter for subblock boundary

In VVC3, deblocking filter is enabled on 8x8 grid as HEVC. For ATMVP and affine sub-blocks on 8x8 grid, the same logic in PU in HEVC deblocking filter is applied. For PU boudaries, HEVC deblocking filtering is applied on 8x8 grid considering following cases:

- If the edge is a transform block edge and there are coefficients in either block.

- If there are no transform coefficients or block boundary is not a transform block boundary then motion conditions are checked (i.e. the difference between motion vectors and reference pictures).

Therefore, HEVC deblocking condition in PU boundaries is taken into ATMVP and affine sub-blocks in VVC. This means check the deblocking motion conditions for ATMVP and affine motion sub-block boundaries as if they were PUs in HEVC.

## 360-degree video coding tools

### Horizontal wrap around motion compensation

The horizontal wrap around motion compensation in the VTM-3.0 is a 360-specific coding tool designed to improve the visual quality of reconstructed 360-degree video in the equi-rectangular (ERP) projection format [5]. In conventional motion compensation, when a motion vector refers to samples beyond the picture boundaries of the reference picture, repetitive padding is applied to derive the values of the out-of-bounds samples by copying from those nearest neighbors on the corresponding picture boundary. For 360-degree video, this method of repetitive padding is not suitable, and could cause visual artefacts called “seam artefacts” in a reconstructed viewport video. Because a 360-degree video is captured on a sphere and inherently has no “boundary,” the reference samples that are out of the boundaries of a reference picture in the projected domain can always be obtained from neighboring samples in the spherical domain. For a general projection format, it may be difficult to derive the corresponding neighboring samples in the spherical domain, because it involves 2D-to-3D and 3D-to-2D coordinate conversion [5], as well as sample interpolation for fractional sample positions. This problem is much simpler for the left and right boundaries of the ERP projection format, as the spherical neighbors outside of the left picture boundary can be obtained from samples inside the right picture boundary, and vice versa. Given the wide usage of the ERP projection format, and the relative ease of implementation, the horizontal wrap around motion compensation was adopted in the VTM-3.0 to improve the visual quality of 360-video coded in the ERP projection format.



**Figure 36 – Horizontal wrap around motion compensation in VVC**

The horizontal wrap around motion compensation process is as depicted in **Figure 36**. When a part of the reference block is outside of the reference picture’s left (or right) boundary in the projected domain, instead of repetitive padding, the “out-of-boundary” part is taken from the corresponding spherical neighbors that are located within the reference picture toward the right (or left) boundary in the projected domain. Repetitive padding is only used for the top and bottom picture boundaries. As depicted in **Figure 36**, the horizontal wrap around motion compensation can be combined with the non-normative padding method often used in 360-degree video coding (see padded ERP in [5]). In VVC, this is achieved by signaling a high level syntax element to indicate the wrap-around offset, which should be set to the ERP picture width before padding; this syntax is used to adjust the position of horizontal wrap around accordingly. This syntax is not affected by the specific amount of padding on the left and right picture boundaries, and therefore naturally supports asymmetric padding of the ERP picture, i.e., when left and right padding are different. The horizontal wrap around motion compensation provides more meaningful information for motion compensation when the reference samples are outside of the reference picture’s left and right boundaries. Under the 360 video CTC [4], this tool improves compression performance not only in terms of rate-distortion performance, but also in terms of reduced seam artefacts and improved subjective quality of the reconstructed 360-degree video. The horizontal wrap around motion compensation can also be used for other single face projection formats with constant sampling density in the horizontal direction, such as adjusted equal-area projection in 360Lib [5].

## Screen content coding tools

### Current picture referencing (CPR)

Current picture referencing is sometimes referred to as intra block copy, where a motion vector refers to the already reconstructed reference samples in the current picture. CPR was supported in HEVC screen content coding extension (HEVC SCC). A CPR-coded CU is signaled as an inter coded block. The luma motion (or block) vector of a CPR-coded CU must be in integer precision. The chroma motion vector is clipped to integer precision as well. When combined with AMVR, the CPR mode can switch between 1-pel and 4-pel motion vector precisions. The current picture is placed at the end of the reference picture list L0. To reduce memory consumption and decoder complexity, the CPR in VTM-3.0 allows only the reconstructed portion of the current CTU to be used. This restriction allows the CPR mode to be implemented using local on-chip memory for hardware implementations.

At the encoder side, hash-based motion estimation is performed for CPR. The encoder performs RD check for blocks with either width or height no larger than 16 luma samples. For non-merge mode, the block vector search is performed using hash-based search first. If hash search does not return valid candidate, block matching based local search will be performed.

In the hash-based search, hash key matching (32-bit CRC) between the current block and a reference block is extended to all allowed block sizes. The hash key calculation for every position in the current picture is based on 4x4 sub-blocks. For the current block of a larger size, a hash key is determined to match that of the reference block when all the hash keys of all 4×4 sub-blocks match the hash keys in the corresponding reference locations. If hash keys of multiple reference blocks are found to match that of the current block, the block vector costs of each matched reference are calculated and the one with the minimum cost is selected.

In block matching search, the search range is set to be N samples to the left and on top of the current block within the current CTU. At the beginning of a CTU, the value of N is initialized to 128 if there is no temporal reference picture, and initialized to 64 if there is at least one temporal reference picture. A hash hit ratio is defined as the percentage of samples in the CTU that found a match using hash-based search. While encoding the current CTU, if the hash hit ratio is below 5%, N is reduced by half.

# Description of VTM3 encoder and encoding methods

### Derivation process of coding tree structure

To be added.

# References

1. B. Bross, J. Chen, S. Liu, “Versatile Video Coding (Draft 3),” document JVET-L1001, 12th JVET meeting: Macao, CN, 3–12 Oct. 2018.
2. F. Bossen, J. Boyce, X. Li, and V. Seregin, K. Sühring, “JVET common test conditions and software reference configurations for SDR video,” document JVET-L1010, 12th JVET meeting: Macao, CN, 3–12 Oct. 2018.
3. A. Segall, E. François, and D. Rusanovskyy, “JVET common test conditions and evaluation procedures for HDR/WCG video,” document JVET-L1011, 12th JVET meeting: Macao, CN, 3–12 Oct. 2018.
4. P. Hanhart, J. Boyce, and K. Choi, “JVET common test conditions and evaluation procedures for 360° video,” document JVET-K1012, 11th JVET meeting: Macao, CN, 3–12 Oct. 2018.
5. Y. Ye, J. Boyce, “Algorithm descriptions of projection format conversion and video quality metrics in 360Lib Version 7,” document JVET-L1004, 12th JVET meeting: Macao, CN, 3–12 Oct. 2018.
6. High Efficiency Video Coding (HEVC), Rec. ITU-T H.265 and ISO/IEC 23008-2, Jan. 2013 (and later editions).
7. G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard”, *IEEE Trans. Circuits and Systems for Video Technology*, Vol. 22, No. 12, pp. 1649‒1668, Dec. 2012.