• RESEARCH PAPER •

Evolution of AVS video coding standards: twenty years of innovation and development

Siwei MA[1*], Li ZHANG[2], Shiqi WANG[3], Chuanmin JIA[1], Shanshe WANG[1], Tiejun HUANG[1], Feng WU[4] & Wen GAO[1]

[1]*Institute of Digital Media, School of Computer Science, Peking University, Beijing 100871, China;*
[2]*Bytedance Inc., San Diego CA 92122, USA;*
[3]*Department of Computer Science, City University of Hong Kong, Hong Kong 999077, China;*
[4]*CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China, Hefei 230027, China*

Abstract Ever since the founding of the Audio and Video Coding Standard (AVS) Workgroup of China in 2002, it has been dedicated to advancing and innovating the digital audio-video industry with highly efficient and economical encoding/decoding technologies. Three representative generations of video coding standards have been finalized and published, consistently improving the coding performance in the past two decades. The series of AVS standards establish solid foundations for ubiquitous video applications in the areas including acquisition, coding, production, delivery, integrated system, public service, general screen content, and mixed reality media. Along with the standardization process, an extensive amount of studies have been carried out on efficiency-aware designation, algorithm optimization, and hardware implementation of these innovative video coding techniques. This paper explains how those developed techniques provide a lasting impact on the video coding community, extensively, technologically, and systematically. In particular, we provide a comprehensive survey of the three generations of the standards, and timely and in-depth summarize the efforts of the AVS video coding standards in the twenty years. The rate-distortion performance comparisons, in particular in terms of the 8K ultra-high-definition (UHD) contents, reflect the elegant design of the state-of-the-art AVS3 standards. We have also elaborated on a variety of well-established and promising applications, including commercial level real-time 8K encoder, high-frame-rate decoder chip for cell phone, and live streaming solution for sports. In addition, the China Central Television (CCTV) of China Media Group (CMG), the state television of China, has officially launched the first 8K broadcasting channel (CCTV-8K) since 2021 using AVS3. Given the significant success realized by the AVS standards, it is envisioned that a new era of 8K UHD video is arriving.

Keywords video coding, AVS standards, video applications, intelligent coding

1 Introduction

Driven by the increasing demands on compressing the exponentially growing visual data, the Audio and Video Coding Standard (AVS) working group has been continuously working on developing efficient video coding standards for the past two decades. Since the establishment of AVS workgroup in March 2002, a series of prestigious video coding standards and extensions, including AVS1 [1], AVS2 [2], and AVS3 [3], have been published and standardized, receiving increasing attention from both academic society and industry entity. The series of AVS standards are renowned for promising coding performance, hardware-friendly design, intelligent technology enabled coding tools, and transparent intellectual property rights (IPR) policy.

The major timeline of AVS series video coding standards is illustrated in Table 1, in which several representative milestones are listed and highlighted. The roadmap of AVS standards typically collaborates

---

* Corresponding author (email: swma@pku.edu.cn)

Table 1   Timeline of AVS video coding standards in the past two decades

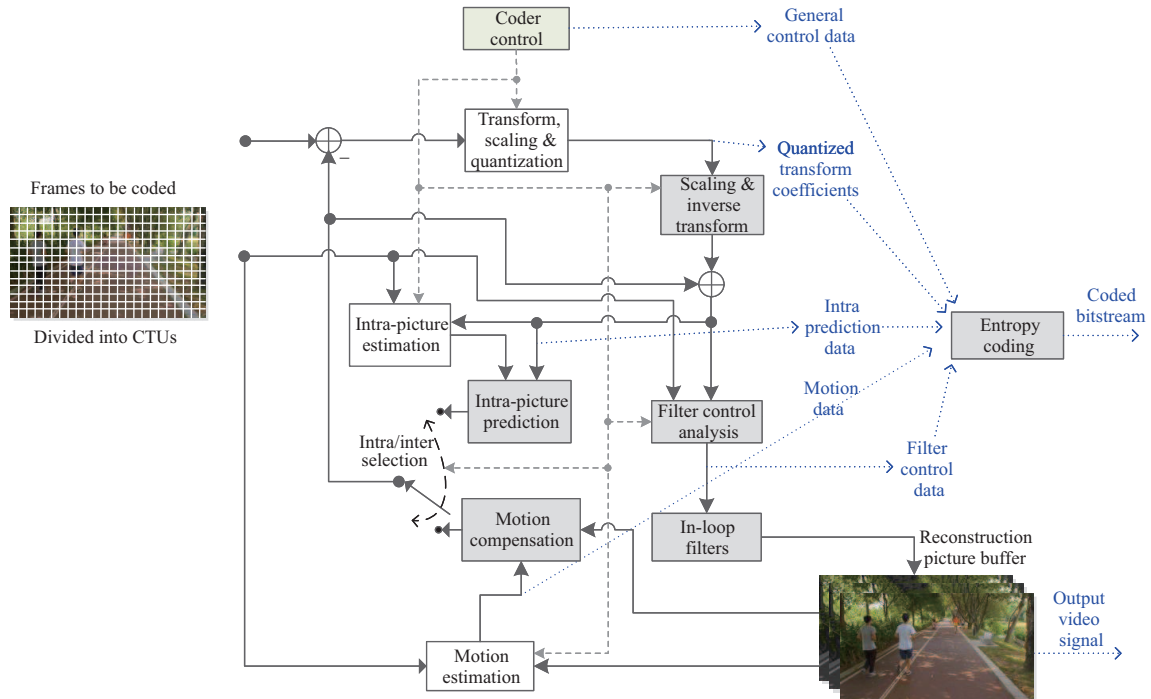| Timeline	Profile	Target application(s)
Dec. 2003	AVS1 Main	Digital TV broadcasting
Jun. 2008	AVS-Surveillance	Surveillance video coding
May. 2012	AVS+	HDTV broadcasting
Dec. 2014	AVS2 Main	4K TV broadcasting
Apr. 2018	AVS2 3D	3DTV, multiview video
Mar. 2019	AVS3 Main (Phase-1)	8K UHD TV broadcasting, VR
Jun. 2021	AVS3 High (Phase-2)	

with the proliferation of video acquisition/display devices and the growing popularity of high quality visual services. The first generation AVS video coding standard (AVS1) was finalized in December 2003. The objective of AVS1 Main (Jizhun) Profile is to provide a light-weight solution for digital standard-definition (SD) TV broadcasting. It was officially approved as a national standard in China in 2006 under grant number GB/T 20090.2-2006. Subsequently, an enhanced version, called AVS+, was established in May 2012 for digital high definition TV (HDTV) broadcasting services, which was a major standard for cable networks. In 2014, AVS working group developed its second-generation video coding standard, AVS2, for 4K ultra-high definition (UHD) TV broadcasting. AVS2 standard made the video coding smarter than ever by supporting various computer vision tasks using content-analysis based compression and scene coding. The background modeling technique in AVS2 significantly reduces the bit-rate when encoding the scene contents such as public service videos. In analogous to its predecessor, AVS2 was also recognized as the Chinese national standard in 2016 under grant number GB/T 33475.2-2016. It has also been approved as a project of IEEE standard, IEEE 1857.4. In addition, two extended profiles, AVS2 multi-view (MV) and AVS3 3D, were developed based on the AVS2 Main Profile for multi-view coding and 3DTV, respectively, for use cases in free viewpoint TV and 3D display.

The major objective of developing a new generation AVS standard in the post-stage of AVS2 is to satisfy the demand for compressing the emerging media data (including UHD contents, 360 video, virtual reality, and user-generated contents), with 4K, 8K, or even higher resolutions. According to the Cisco annual internet report[1], the total number of Internet users is projected to grow from 3.9 billion in 2018 to 5.3 billion by 2023. Moreover, the major of network traffic will be occupied by UHD video data. This implies that the efficiency of UHD video compression is critical for the quality-of-experience in videos watched by half of the global population. Moreover, during the COVID-19 pandemic, the video conferencing and on-line video services have raised higher demand for compression technologies to ensure the quality-of-experiences. To this end, the AVS working group officially initiated the next generation video coding standard AVS3 in March 2018, aiming at achieving significant bit-rate saving than any other preceding AVS standards and in harmony with both software and hardware implementation.

The development of AVS3 standard follows a progressive two-step strategy approved by the AVS working group. In particular, there are two phases for AVS3. The reference software and text specification of AVS3 Main Profile (Phase-1) have been finalized in March 2019. There are a variety of novel coding tools adopted in AVS3 Main Profile while the encoding and decoding complexity is well-controlled to balance the performance complexity trade-off. After two years of development, the AVS3 High Profile (Phase-2) have been accomplished in June 2021, which officially announced that the state-of-the-art AVS3 standard is finalized and standardized. Successive exploration is also being conducted to enhance the coding performance for the preparation of the subsequent standards, which further improves the compression efficiency.

This article provides an overview of the innovation history of the series of AVS video coding standards associated with the innovative technologies. The remainder of this paper is organized as follows. Section 2 reviews the advancement of the coding techniques from AVS1 to AVS3. Section 3 elaborates the development of AVS series standards from scratch with an emphasized technical roadmap description on AVS3. The compression efficiency of different generations of AVS standards is quantitatively analyzed in Section 4. In Section 5, the real-world applications and deployment of AVS3 standards are extensively introduced, and Section 6 concludes this paper.

---

1) https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

Figure 1 (Color online) Illustration of the coding framework of the series of AVS standards. Gray modules indicate the procedures that are decoder-only.

2 Key technologies and features in AVS standards

The series of AVS standards adopt the classical hybrid block-based predictive-transform coding framework. The framework is shown in Figure 1, which mainly consists of the following major modules, prediction (including intra and inter prediction), transform and quantization, loop filter, and entropy coding. Each coding module has been elaborately designed and enhanced for further promoting the coding efficiency. This section provides a brief introduction to the framework and key features of the series of AVS standards.

2.1 Block structure

In the classic hybrid-coding framework, video coding is operated in the unit of macroblocks after dividing each frame of the video sequence into invariable-sized macroblocks. Generally speaking, the block with smaller size can provide more accurate prediction both for intra- and inter- frames, corresponding to smaller prediction residual and improved prediction efficiency. However, more bits need to be consumed to transmit the extra motion information and the intra prediction mode. Experiments show that the performance of macroblocks with fixed size 8×8 is better than that of macroblocks with fixed size 4×4. Therefore, the size of macroblocks is fixed to 8×8 in AVS1. For inter prediction, four types of inter-macroblock partitions have been applied to motion compensation in AVS1. As illustrated in Figure 2, the luminance component for each inter-macroblock allows one 16×16 block, two 16×8 blocks, two 8×16 blocks or four 8×8 blocks for motion compensation. No further partition for an 8×8 block is allowed considering that high-resolution videos usually have strong spatial correlation among the neighboring pixels within a picture, and 2D 8×8 discrete cosine transform (DCT)-like integer transform is used correspondingly. In addition, intra prediction is carried out for 8×8 luminance and chroma blocks.

Constrained by the fixed block sizes and shape, the partition schemes based on macroblocks cannot cope with the demand arising from video content with larger resolutions. Therefore, AVS2 adopts a flexible nested quad-tree based partition structure. In particular, pictures are firstly split into largest coding units (LCUs) with a fixed size, and LCUs can be recursively divided into coding units (CUs) of different sizes by quad-tree partition, followed by the subsequent prediction and transform coding. The LCU size up to 64×64 makes the size of CUs vary from 8×8 to 64×64, which enhances the adaptability of the structure on different scales. To further increase the flexibility of partition, asymmetric partition
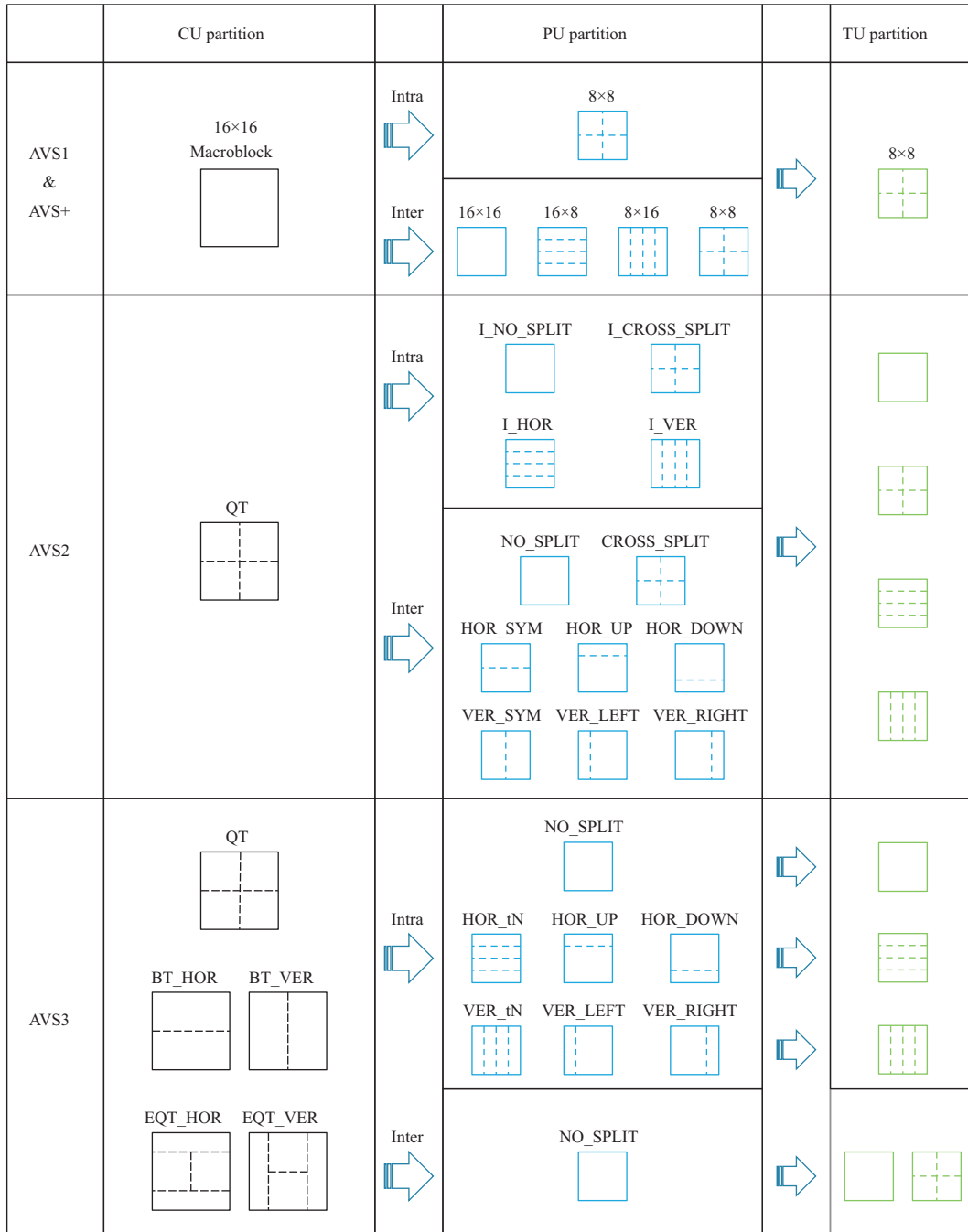
Figure 2   (Color online) Evolution of block partition schemes of AVS standards. The partition structure becomes more complicated and content-adaptive along with the development of the standards.

methods are introduced for prediction coding and transform coding. In AVS2, prediction coding is based on prediction units (PUs) partitioned from CUs. For intra prediction, short distance intra prediction (SDIP) is used to divide a $2N \times 2N$ CU into four PUs with an aspect ratio of 4. Such partition persuades nearer reference pixels to be utilized to generate more accurate prediction results. For inter prediction, asymmetric motion partition (AMP) adds extra four partition methods to benefit motion estimation, which can partition a CU into four PUs with the same size or two PUs with different sizes. In analogous to prediction, transform coding in AVS2 is performed in transform units (TUs) split from CUs. Considering that performing transform with the combination of multiple residual blocks may increase the bitrate of transform coefficients, non-square quadtree transform (NSQT) is added to split a CU into four TUs according to PU partition in the current CU.

AVS3 supports more flexible coding tree and block partitioning schemes to adapt the diverse content of textures in finer-granularity and growing video resolutions. The coding tree unit (CTU) size in AVS3 can reach up to 128×128 and be partitioned into the smallest 4×4 blocks. Besides the quad-tree and binary tree (QTBT) partitioning, the extended quad-tree partition (EQT) is developed for further improving the prediction accuracy [4]. Multiple types of block partition schemes are adopted in AVS3, including quad tree (QT), binary tree (BT) partition, and EQT partition, which are shown in Figure 2. It is worth mentioning that there is an additional derived tree (DT) [5] partitioning for intra coding. The CTU is further partitioned into CUs through different partition modes. For CUs, the size and mode for PU and TU are subsequently determined in the sense of RD optimization.

2.2 Intra prediction

The intra prediction modes of AVS are composed of DC mode, plane mode, bi-linear mode, and angular modes. Angular modes can capture the edge directions in natural videos while other modes are applicable to flatten and gradual texture prediction. AVS1 supports up to eight angular intra modes as illustrated in Figure 3(a), and the usage of these modes varies with block sizes [6]. All the angular modes are available for 4×4 intra prediction since it leads to better coding efficiency. For 8×8 block, only the vertical, horizontal, and diagonal modes can be used for prediction [7]. For the vertical and horizontal prediction modes, the above or left neighboring samples are directly used as the prediction samples for the corresponding entire column or row. Meanwhile, 8×8 intra prediction allows two special modes, DC mode, and Plane mode, for the luminance and chrominance components, respectively. In DC mode, each prediction sample is obtained by an average of the corresponding vertical and horizontal reference pixels. Hence, the prediction values of different pixels in a block might be different [8]. To efficiently encode the multiple intra predicition modes, the two most probable modes (MPM) are proposed for intra mode coding. The MPMs are estimated based on the spatial correlations between the current PU and the adjacent PUs directly. The first MPM is the mode with smaller index from the left PU and the upper PU, and the other is the second MPM. As such, only 1-bit is used to code the MPM, consuming fewer bits compared to fixed length coding (FLC) for regular modes.

Owing to the increasing PU sizes, the nine prediction directions in AVS1 cannot accurately handle the complicated textures. To significantly improve the prediction accuracy, the angular modes with finer angularity were adopted in AVS2 [9]. As shown in Figure 3(b), up to 33 angular modes can be used for each PU, including 30 angular modes and 3 special modes. The prediction directions associated with the 30 angular modes are distributed within the range of $[-157.5°, 60°]$. For modes from 3 to 11, only the top neighboring samples are used for prediction, and for modes from 25 to 32, only left neighboring samples are used for prediction. For modes from 13 to 23, top or left neighboring samples are employed for prediction according to the intersections between the directional lines and reference boundaries. Different from the DC mode in AVS1, the DC mode of AVS2 considers all available reference samples for averaging and uses the average as the prediction sample for the whole block. Besides, in AVS2, a special mode termed Bilinear mode was adopted [10]. The bilinear mode uses the top, left, bottom left, and top right pixels to generate prediction samples through two interpolations and it is suitable for gradually changing areas. Owing to the introduction of Bilinear mode, minor changes are made to the construction of MPM list [11]. If both of the two neighboring blocks use DC mode, the DC and bilinear mode are used as the two MPMs. Otherwise, DC mode and the prediction mode of the neighboring blocks are derived as the two MPMs.

AVS3 inherits the DC, plane, and bilinear modes from AVS2 and the range of angular directions remains unchanged. Moreover, the number of angular prediction directions has been increased [12]. As shown in Figure 3(b), denser angular modes are inserted equally spaced based on the horizontal or vertical projection position. Since the vertical and horizontal modes are adopted more frequently, more angular modes are inserted around the two modes to improve the efficiency of entropy coding. The generation of MPM list for the 66 intra modes follows AVS2. To better predict complex textures with multiple directions, the spatial angular weighted prediction (SAWP) method is proposed in AVS3, where weighted prediction is achieved by the binary combination of existing uni-intra prediction modes [13]. More specifically, two prediction modes are leveraged to generate the prediction blocks. Then the final prediction is generated by the weighted combination of the two prediction blocks and the weights are derived by the specific templates. When SAWP is used, two prediction modes instead of one prediction mode are transmitted in the bitstream. As such, only mode 3 to mode 32 are available for SAWP to
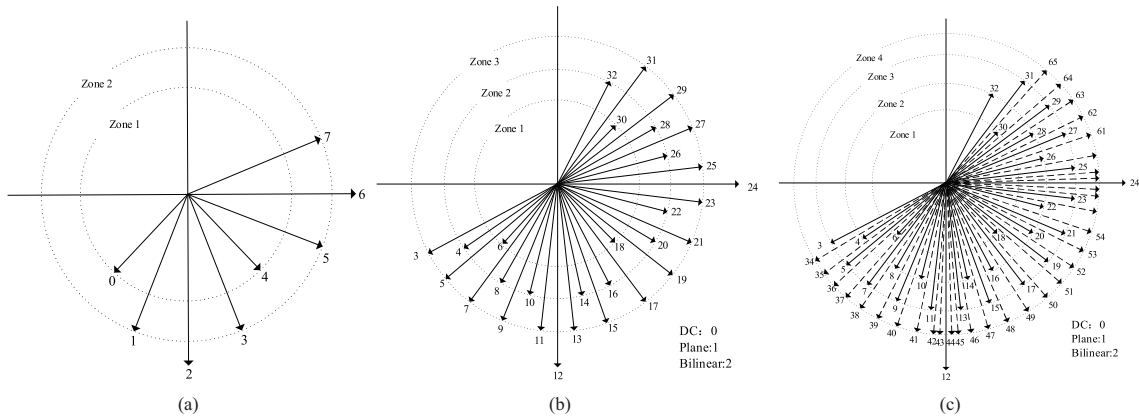
Figure 3 The intra prediction modes of AVS video coding standards. (a) 8 intra prediction modes in AVS1; (b) 33 intra prediction modes in AVS2; (c) 65 intra prediction modes in AVS3.

reduce the signaling overhead.

Given an angular mode, each sample in a PU is predicted by projecting its location to the reference boundaries and the corresponding reference pixels are used as prediction values. In AVS1, only the integer positions have been considered for prediction and a three-tap low-pass filter (1, 2, 1) is applied for DC mode and diagonal modes. With the increase of angular prediction modes, the sub-pixel positions cannot be ignored. To improve the prediction accuracy, the reference samples of sub-pixel precision are interpolated using adjacent integer samples in AVS2. All the non-integer positions between two integer positions are aligned to 1/32 sample precision and a four-tap linear interpolation filter is used to generate the fractional reference samples [14]. To further eliminate the noise and obtain more accurate references when the angular direction points to a fractional position, multiple intra prediction filter (MIPF) in AVS3 adopts four types of four-tap filters [15]. The multiple filters are composed of smoothing filters and interpolation filters and each filter has different degree of smoothness. The smoothing filter flattens the values of a prediction block, making it transform friendly due to the removal of the detailed information. Hence the filter type is alternatingly chosen according to the distance from the reference samples and the intra prediction mode. The four-tap filters are used in AVS2, as multiple filters which have different degrees of smoothness in the same block can improve both accuracy and naturalness of the intra prediction block.

Different from AVS1 and AVS2, the prediction results of intra modes can be further modified by intra prediction filter (IPF) [16]. IPF refines the intra prediction results by using the neighboring reference samples, aimed at improving the prediction accuracy. Besides exploring the correlations within the same component in AVS1 and AVS2, the relationship between the luma channel and chroma channel has also been comprehensively investigated in AVS3. The correlations among different color components were explored in the early explorations focusing on the linear relationship between different components of RGB 4:4:4 color space. The linear model can be implemented with the explicit solution, where linear parameters are signaled from the encoder, as well as the implicit solution, where linear parameters are derived at the decode. To reduce the cross-component redundancy, AVS3 adopts two step cross-component prediction mode (TSCPM), which reconstructs the chroma samples by reconstructed luma sample. The linear model is designed as $C = \alpha \times L + \beta$, where C is the chroma samples to be reconstructed, L denotes the already reconstructed luma samples. α and β are two linear parameters, which are calculated according to [17]. TSCPM reveals great benefits in promoting the coding efficiency by removing the inter-channel redundancies between the luma component and chroma component. To further remove the inter-channel redundancy, another chroma prediction mode, named prediction from multiple cross-components (PMC) [18], was proposed. With the PMC mode, the Cr coding block can be predicted through a linear combination of the reconstructed coding blocks regarding Y and Cb components. The chroma coding tools, including TSCPM and PMC, efficiently remove the inter-channel redundancies and achieve remarkable coding gains.

2.3 Inter prediction

Inter prediction has been an essential module since AVS1 due to its capability in eliminating temporal redundancy. During the development of AVS video coding standards, a variety of new coding tools for more efficient prediction and coding of motion information, as well as for enhancing the motion compensation, have been comprehensively explored. These techniques can be categorized into: (a) predictive coding types; (b) motion information coding; (c) CU-level and subblock-level motion compensation. These three categories are detailed as follows.

2.3.1 *Developments of predictive coding types*

During the evolution of AVS video coding standards, multiple predictive frame types have been considered. In AVS1, there are two different picture types according to the temporal prediction: P frame corresponds to previous-prediction (p-prediction) and B frame corresponds to bi-prediction. P-prediction restricts the reference of prediction to the decoded pictures which are before the current coding picture in display order. Bi-prediction stands for the inter-frame prediction from the forward- and backward-decoded pictures in display order. In AVS2, a new inter frame type, F frame, is defined as a special P frame, which enables the prediction from two forward references [19]. In AVS3, the concept of B frame is generalized. The generalized B frame removes the restriction of the bi-prediction to allow only linear combinations of forward and backward pairs. In other words, bi-predictive block can use two prediction blocks from an arbitrary set of reference pictures in forward and/or backward prediction directions.

In AVS1, there are four possible coding modes for each inter marcoblock (MB): inter-, skip-, direct- and symmetric-MBs. Direct mode and symmetric mode are two unique techniques of bi-prediction in AVS1 [20]. In direct mode, both forward and backward MVs of current block are derived from the MV of its collocated block in the backward reference frame according to the temporal distance between predicted and reference blocks. While in symmetric mode, forward MV needs to be transmitted, while backward MV is derived from the forward MV using a symmetric rule. In AVS2, inter prediction mode has been improved with the use of multi-hypothesis techniques, including multi-directional skip/direct mode, temporal multi-hypothesis prediction mode [21], and spatial directional multi-hypothesis (DMH) prediction mode [22].

2.3.2 *Developments of motion information coding*

Motion information coding has been a key technique in inter-prediction since AVS1, including motion information prediction and motion vector difference (MVD) signaling. Motion information prediction can reduce the redundancy among MVs of spatially/temporally neighboring blocks and thus save a large number of bits for MV coding. In AVS1, MVs of spatially neighboring blocks and temporal collocated blocks were used for the motion information prediction. In AVS2, the motion information prediction methods were extended using median MV prediction and spatio-temporal MV prediction. In AVS3, these methods were further improved by advanced predictors, such as motion vector angle prediction (MVAP) and history-based MV prediction (HMVP). MVD signaling method is mainly determined by the MV resolution. With the evolution of AVS standards, more available MV resolutions are supported to improve the trade-off between motion accuracy and motion overhead bits.

• Median MV prediction. Median MV prediction aims to generate a new MV predictor with the MVs of three spatial neighboring blocks, i.e., the up, left, and up-right blocks. The new MV predictor is calculated by averaging the two relatively similar MVs.

• MVAP. The MVAP technique is an 8×8 subblock-based MC method designed for direct/skip mode in AVS3 [23]. As depicted in Figure 4, there are five angle candidates designed in MVAP. Given the prediction angle predicted from the corresponding neighboring blocks, the MV of each subblock can be obtained.

• HMVP. The motivation of HMVP [24] is to explore the potential of history information for better motion information prediction. The HMVP candidates are recorded using a table, which is updated with the motion information of the previously coded blocks based on a first-in-first-out (FIFO) rule. As such, the HMVP candidates serve as additional modes for skip or direct modes to improve the coding efficiency.

• Progressive motion vector resolution (PMVR). In AVS1, the resolution of MV is fixed to 1/4-pixel. To further reduce coding bits of MVD, PMVR is introduced in AVS2 [25]. In PMVR, the MV resolution is progressively adjusted based on the distance between the MV and motion vector prediction (MVP),
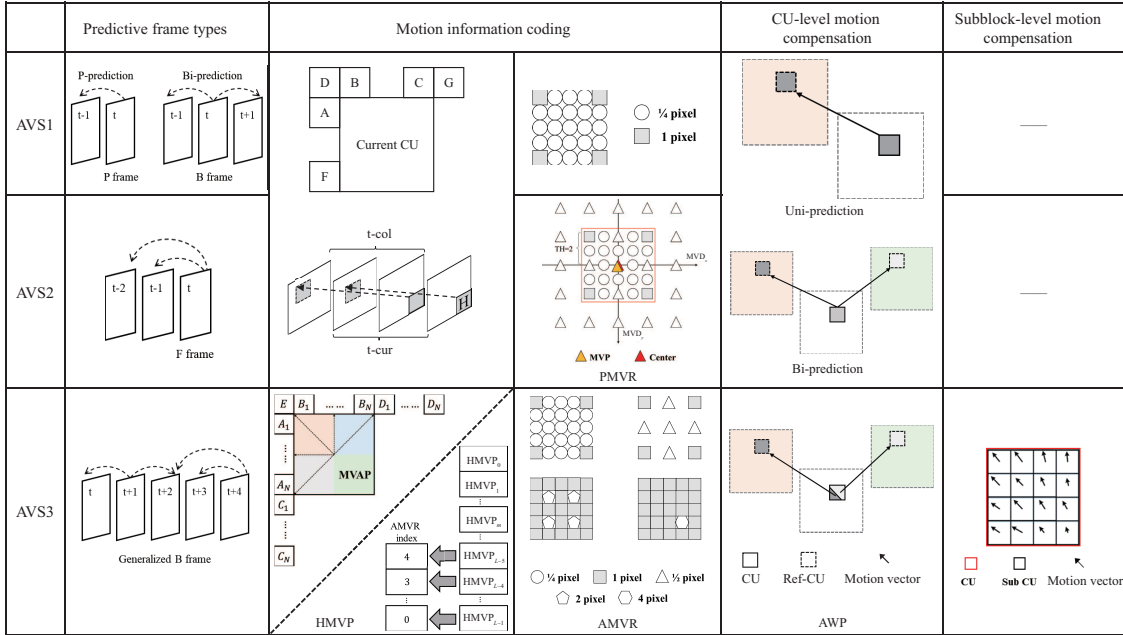
Figure 4 (Color online) The comparative technical development of inter prediction tools in AVS standards.

and the ranges of resolution are restricted by a simple threshold. According to the observation that the MV closer to the MVP is more likely to be optimal in the rate distortion sense, higher MV resolution is employed for MVs near to the MV and lower MV resolution is used for MVs far from the MVP. More specifically, the 1/4-pixel MV positions are disabled when the MVs are outside of the specific 1/4-pixel range. Figure 4 illustrates the MV resolution restriction, in which the red square indicates the 1/4-pixel range.

• Adaptive MV resolution (AMVR). AVS3 supports AMVR at the CU level to allow a better trade-off between MV overhead and prediction accuracy. For inter-predicted CUs with translation motion mode, the MV resolutions can be selected from 1/4-pixel, 1/2-pixel, 1-pixel, 2-pixel, and 4-pixel. For Affine-coded CUs (see Subsection 2.3.3), 3 MV resolutions are supported, i.e., 1/16-pixel, 1/4-pixel, and 1-pixel.

2.3.3 *Developments of motion compensation*

Motion-compensated prediction is widely used in video coding process to eliminate the temporal redundancy. Given the forward and/or backward frames, the current frame can be predicted using the prior knowledge of motion models. With the evolution of AVS standard, motion compensation technologies have been enhanced and gradually refined.

Derivation of prediction sample at CU-level. Block-based motion compensation is applied since AVS1 and still plays an important role in AVS3. In AVS1 and AVS2, reconstructed samples of reference blocks are used as prediction samples directly. In AVS3, more enhanced MC tools are introduced, including prediction for oblique boundary regions inside one CU and combination of multiple prediction signals.

• Angular weighted prediction (AWP). AWP is designed for the coding of oblique boundary regions of two objects [26]. It conducts angular weighted prediction using the two predicted blocks of each CU. To meet this demand, eight angles are supported in AWP. The weights in the AWP are selected from the pre-defined reference weight sets. Given the angle and reference weight set, each sample can derive its corresponding weight, as shown in Figure 4.

• Overlapped block motion compensation (OBMC). The OBMC technique is introduced to improve the prediction quality in AVS3 [27]. Using the MV of the current block and the MV of its neighboring blocks (i.e., top and left neighbors), multiple prediction samples can be obtained. Subsequently, the prediction samples along the boundaries of the current block and its casual neighbors are blended to generate the corresponding OBMC prediction signal through weighted average operations. In addition, the OBMC is only enabled for inter CUs that are uni-predicted, and is always disabled for the small CUs

(e.g., smaller than 64 samples) and the CUs that are coded by AWP mode.

• Inter prediction filter. To alleviate the discontinuity between the prediction sample and its neighboring samples, an inter prediction filter [28] is introduced in AVS3 to refine the prediction signal. For CUs coded with direct mode, the filtered prediction samples are generated by the weighted averaging process. The input to this filtering process includes the prediction sample and its neighboring samples, i.e., upper and left, while the weighting factors are determined according to the relative coordinates of each sample.

Derivation of prediction sample at subblock-level. In AVS1 and AVS2, motion compensation is performed at marcoblock/CU level. Using fine-granular motion representation can further improve the prediction accuracy and help exploiting inter-frame correlation in video coding. Therefore, AVS3 enhances inter prediction by introducing technologies that obtain fine-granular motion information. In addition, the prediction samples can be further refined by optical flow-based coding tools. These tools are described as follows.
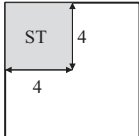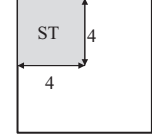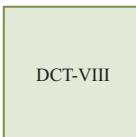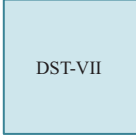
• Affine motion compensation (AMC). In AVS1 and AVS2, only translation motion model is applied in motion compensation. In real world scenarios, there exist various motions such as rotation and zooming. In order to address this issue, a block-based affine MC is introduced in AVS3 [29]. The affine motion can be represented with a 4-parameter model or a 6-parameter model. To reduce the complexity of AMC, the MV granularity in AVS3 is at the subblock level instead of pixel level. In order to derive MV of each subblock, the MV at the center point is calculated by motion information at two control points with the 4-parameter affine motion model. AMC is applied to generate the prediction samples of each subblock with derived MVs. In analogous to the translational motion model, there are two affine motion prediction modes: affine inter and affine direct mode.

• Decoder-side motion vector refinement (DMVR). In AVS3, in order to reduce the overhead of transmitting motion parameters, DMVR is applied to derive motion parameters at the decoder side [30]. Based upon the pair of MVs derived from direct/skip mode as initial MVs, DMVR applies bilateral matching to refine the initial MVs. The refined MVs are obtained in two steps, an integer-sample search step within the search range of ± 2 integer luma samples followed by a fractional-sample search step. For each subblock, the corresponding refined MV pairs are used for motion compensation.

• Bi-directional optical flow (BIO). To compensate the sample-wise motion that is missed by the block-based motion compensation, BIO is applied in AVS3 as another decoder-side prediction tool [31]. According to the optical flow differential equation which minimizes the difference between the prediction subblocks in forward and backward reference frames, the motion refinements are derived implicitly from the samples of two prediction blocks for each 4×4 subblock. For feasible hardware implementations, BIO is only applied for CUs smaller than 64×64 and all the BIO related computations can be implemented using integer arithmetic not exceeding 32 bits.

2.4 Transform coding and quantization

Transform. The development history of transform coding techniques in a series of AVS standards is summarized in Figure 5. Transform coding tools have been optimized in a coarse-to-fine manner. In AVS1, integer DCT is applied to prediction residuals, as integer DCT is computational-efficient and has better properties in encoder-decoder match compared with floating-point DCT. The baseline profile only applies integer DCT of 8×8 block size. In other profiles, the transform block size is extended to 4×4 and 16×16 blocks. For 8×8 blocks, 8×8 transform and 4×4 transform can be adaptively applied by signaling an indicator in the bitstream. AVS2 extends the block size of TU, and the maximum size of TU is 64×64. For a 64×64 TU, a 5-3 tap integer wavelet transform is performed to reduce the TU size to 32×32, and then DCT is applied. Rectangular DCT is applied in AVS2, and the size is from 4×4 to 32×32. Moreover, for CUs with asymmetric PU partition, NSQT is applied to decrease the value of high-frequency coefficients caused by the increase of residuals on PU boundaries. Besides, the secondary transform is applied for the lowest-frequency 4×4 subblock of DCT transform coefficients of intra-coded residuals. There is a correspondence between the horizontal or/and vertical secondary transform and the intra prediction mode index. The transform module in AVS3 has been optimized from three major perspectives, the position of transform, the transform block granularity, and the coefficients parity. The position-based transform (PBT) introduces a new transform method that uses pre-designed transform sets for four subblocks according to their positions. For each CU, a syntax element PBT flag will be signaled. Regarding the transform block granularity, subblock-based transform (SBT) is proposed to

	Direct transform		Secondary transform	Subblock transform	
AVS1	8×8	DCT-II	—	—	
AVS2	4×4, 8×8, 16×16, 32×32 (64×64 wavelet LL subblock)	DCT-II	Luma — ST 4 / 4	—	
AVS3	4×4, 8×8, 16×16, 32×32, 64×64	DCT-II	Luma, chroma — ST 4 / 4	4×4, 8×8, 16×16 (PBT, SBT)	DCT-VIII
	4×4, 8×8, 16×16, 32×32 (IST)	DST-VII		4×4, 8×8, 16×16 (PBT, SBT)	DST-VII

Figure 5 (Color online) Transform schemes of AVS standards. The transform type becomes more complicated and residual-adaptive along with the development of the standards, yielding higher and higher transform coding efficiency.

capture the partial prediction residual in inter prediction in which residual occupies a part of the current block. In this case, only the part of block is transformed with pre-designed transform core. And the remainder transform coefficients of this block are regarded as all zeros. There are 8 types of transform combination in SBT. Besides only transforming the dominant residual of inter block, the horizontal and vertical transform of the dominant residual region can be selected adaptively to obtain better energy compaction.

Towards saving the transform overhead via the parity of transform coefficients, implicit selected transform (IST) is applied to CUs whose sizes are from 4×4 to 32×32 for intra-coded CUs. One constraint is that Intra-DT partitioned blocks will skip the usage of IST. Two existing transform cores, DCT-II and DST-VII, are employed in IST for the horizontal and vertical transform, respectively. The selection of one of the two transform cores is implicitly determined by the parity of the number of non-zero coefficients. Specifically, the parity of the number of coefficients in one block is employed to represent the transform types. Odd number indicates that the DST-VII is applied, while even number indicates DCT-II is applied.

Since the secondary transform selected by mode-dependent method is not always optimal, enhanced secondary transform (EST) is proposed to indicate whether ST is applied by explicitly signaling a flag. It is applied to non-DT intra-coded CUs, and the primary transform needs to be DCT. More specifically, '0' indicates not using ST, and '1' indicates using ST. Secondary transform for chroma components (ST_CHROMA) extends the EST method to chroma components of CUs.

Quantization. In AVS coding standards, quantization is combined together with the normalization of the transform and is implemented by multiplication and right shift. The quantization parameter is from 0 to 79, and QP 64 to 79 is only used for 10-bit encoding. The corresponding quantization step varies from 1 to 256, allowing for a wide range of compression ratio and application scenarios. The mapping of QP values to quantization step sizes is approximately logarithmic and an increase by 8 of QP doubles the quantization step size approximately. In AVS1, a picture level "QP Shift" parameter can be passed to adjust the quantization step for 4×4 transform coefficients [32, 33]. To provide fine-grained quantization

for different residuals, weighted quantization is adopted and an adaptive quantization mode including three kinds of frequency-weighted quantization modes is proposed in AVS2 and AVS3 [34]. Besides the default weighted quantization parameters, new parameters can be transmitted in the sequence header, or updated in the picture header by transmitting the difference under adaptive quantization mode. For each kind of frequency-weighted quantization mode in adaptive quantization mode, six parameters are used to define the weighted quantization matrix pattern for 4×4 and 8×8 blocks. These one-dimension weighted parameters are mapped to the two-dimension weighted quantization matrix $W(i,j)$ according to three predefined matrix patterns for 4×4 and 8×8 blocks.

2.5 Entropy coding

In AVS1, most syntax elements are encoded with fixed length coding or Exp-Golomb coding. For coefficients transform, a memory efficient method called context-based 2D variable length coding (C2DVLC) [35] is specified. C2DVLC defines multiple 2D-VLC tables to convert run-level pairs to code numbers and use Exp-Colomb codes of code numbers as final code words. The 2D-VLC tables are constructed according to the joint probability of run-level pairs. To decrease memory requirement and increase coding efficiency, only the most frequently occurring run-level pairs are included in the tables. Therefore, through switching 2D-VLC table according to context, C2DVCL can adapt to different content. For those uncovered run-level pairs, an escape coding method is used, which predicts level according to current 2D-VLC table and applies Exp-Golomb coding on the run and the prediction error of level.

Compared to variable length coding methods, arithmetic coding methods have higher coding efficiency. Therefore, an enhanced entropy coding method called context-based binary arithmetic coding (CBAC) [36] is introduced as an optional entropy coding method in the enhanced version of AVS. Different from C2DVLC, CBAC is used to encode not only transform coefficients but also some other syntax elements, such as prediction modes and motion information. It constructs several context models for each syntax element according to historically encoded elements to estimate the probability of next syntax elements. To achieve a more accurate symbol probability prediction, a context weighting technique is applied in coefficients coding. Moreover, by mapping computation from original domain to logarithmic domain, CBAC can use additions to replace multiplications or table look-up operation in common arithmetic coding algorithms when updating the probability estimation. This further reduces the computational complexity of CBAC. In AVS1, only 8×8 and 4×4 blocks are supported in transform coding. Accordingly, coefficients coding is performed on each coefficient block and run-level pairs are generated by zig-zag scan on that. For AVS2, larger block sizes, such as 16×16 and 32×32 are adopted. It is inefficient to inherit CBAC from previous standards directly because of the sparse nonzero coefficient in large block. To improve the coding efficiency, a two-level coefficient coding method based on CBAC [37] was proposed in AVS2. At the first level, coefficient blocks are further partitioned into 4×4 subblocks called coefficient group (CG). A reverse zig-zag scan is performed on all CGs in a coefficient block to locate the last nonzero CG and the position is encoded with CBAC. Then, the CGs before the last CG are encoded in scan order. The second level is encoding the 16 coefficients in each CG.

The optimization for entropy coding in AVS3 is mainly focusing on coefficient coding. The major improvement is to introduce a new coefficient coding method named scan region-based coefficient coding (SRCC) [38]. Specifically, the region with non-zero coefficient is represented by the scan region with spatial index (i.e, SRx, SRy), as shown in Figure 6, where SRx is the x-axis of the right most nonzero coefficient position and SRy is the y-axis of the bottom most non-zero coefficient position. Only coefficients in scan region have non-zero value and are coded from the right-bottom corer of the scan region in the inverse zigzag scan order, instead of CG of 4×4 subblock. The associated context models are also carefully designed according to the position and value of coefficients.

2.6 In-loop filtering

Owing to the block-wise operation and coarse quantization, blocking and ringing artifacts have been inevitably induced in the compressed frames, which significantly degrade the objective and subjective quality. To suppress these compression artifacts, in-loop filtering algorithms were comprehensively explored during the development of AVS standards. Deblocking filter (DBF) was first introduced to remove the blocking artifacts in AVS1 and adjusted in the subsequent AVS standards. To alleviate ringing artifacts, sample adaptive offset (SAO) was applied after the DBF since AVS2. To further improve the quality of reconstructed video signal, adaptive loop filter (ALF) was also adopted by AVS2 as the third
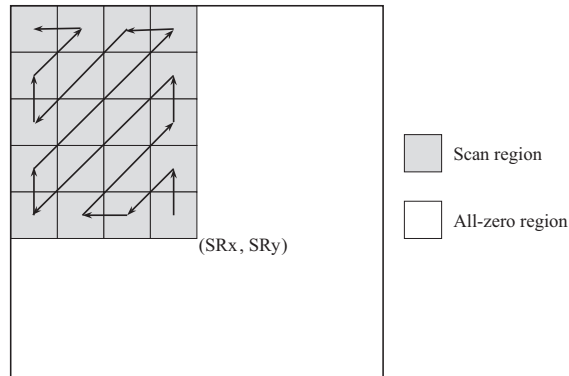
Figure 6 The illustration of SRCC. The white region represents the all-zero coefficients while the gray one shows the scan region with the scanning order for the non-zero coefficients.

in-loop filter after SAO. In AVS3, enhanced sample adaptive offset (ESAO) [39] was contributed as a replacement and improvement of SAO. Besides, cross-component sample adaptive offset (CCSAO) [40] was applied after SAO/ESAO and before ALF, yielding better quality of chroma components.

• **DBF.** The basic unit for the DBF is an 8×8 block. For each 8×8 block, the DBF is used only if the boundary belongs to either of the CU, PU, or TU boundaries. In AVS1, the boundary strength (BS) is dependent on the coding type, quantization step, and motion vectors. Unlike AVS1, the gradient is considered for BS calculation in AVS2, and the number of BS levels is extended from 3 to 5 for better adaptability. In AVS3, the output of DBF is refined based on the encoder-selected parameters to reduce the over-filtering problem. Furthermore, the set of filters is extended by longer filters to reduce artifacts in relatively smooth areas of larger blocks. Compared to the 6-tap filter in AVS1 and AVS2, the DBF in AVS3 enables an 8-tap filter.

• **SAO.** SAO is applied to reduce the mean sample distortion of a region by adding an offset to the reconstructed samples after DBF. In SAO, reconstructed samples are first classified into different categories. Then an offset is obtained for each category. The SAO filtering process is to add the offset to each sample of its corresponding category. There are two modes in SAO: edge offset (EO) mode and band offset (BO) mode. For the EO mode, the sample classification is based on the comparison between the current sample and its neighboring samples. The neighboring samples are derived based on the directional patterns. For the BO mode, the sample classification is based on the amplitudes of the current reconstructed samples.

• **ESAO.** Compared to SAO, ESAO introduces more adaptive sample classification methods which fully consider the textural and edge directional features. There are two feature descriptors in the sample classification process. The first one is calculated based on local binary patterns by comparing the intensities of the current sample with its 8 neighboring samples. The second descriptor is derived based on the amplitudes of the current reconstructed samples. The product of the two feature descriptors leads to the classification result of each sample.

• **CCSAO.** CCSAO is only applied for chroma components. Unlike SAO and ESAO in which reconstructed chroma samples are directly used for the classification, CCSAO classifies the chroma sample utilizing itself and its collocated luma samples. There are two feature descriptors in CCSAO classification process. The first one is based on the difference between the corresponding luma sample and its neighboring samples. The neighboring sample derivation is the same as that of EO mode in SAO. The second descriptor is based on the amplitude of the current sample or its corresponding luma sample, which is determined by the encoder. The final classification result of each chroma sample is calculated by the product of the two descriptors.

• **ALF.** ALF aims to improve the quality of reconstructed video signal by applying a spatial filtering process. It trains filter coefficients in the encoder by using the reconstructed samples after SAO and the original samples according to the principle of minimizing the mean square error (MSE). Then the filter coefficients derived in the encoder are transmitted to the decoder. In AVS2, the spatial filter has 9 coefficients and 17 taps, which is shown in Figure 7(a). In AVS3, a larger filter shape is enabled with 15 coefficients and 29 taps, as shown in Figure 7(b) [41]. To make the filter more adaptive, the reconstructed frame is divided into 16 categories in AVS2, and 16 or 64 categories in AVS3. Filter coefficients are trained for each category. To reduce the heavy burden caused by the coefficient signaling, these filters are merged
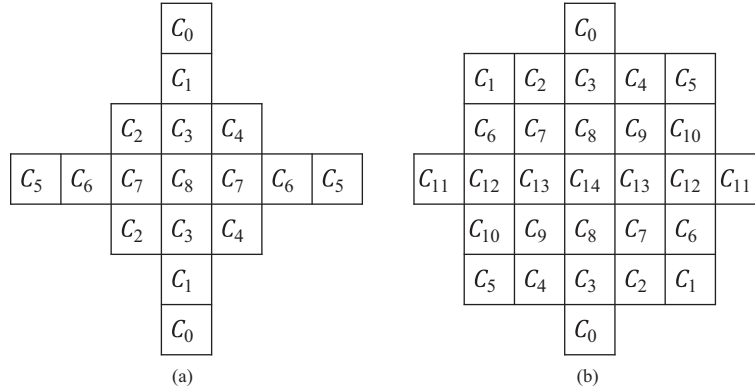
Figure 7 ALF filter shapes in (a) AVS3 Main Profile (also used in AVS2) and (b) AVS3 High Profile.

in the encoder based on the rate-distortion selection.

3 Technical roadmap of the emerging AVS3 standard

This section describes the technical roadmap for the development of AVS3 standard. Generally speaking, the driving force of modern video coding standards is the storage and transmission requirements and emerging use cases. To satisfy various coding requirements and applications in the past two decades, including higher resolution broadcasting services, interactive low delay scenarios, on-demand services or professional video applications, the AVS working group dedicated itself in designing and proposing efficient coding tools on the top of the classical block-based hybrid video coding framework.

The AVS3 standard was primarily initiated in March 2018 by AVS working group, the objective of which is to realize significant coding gain against its predecessors while maintaining better complexity performance trade-off to facilitate diverse use cases. Regarding the main profile, several high efficient coding tools including block partitioning and predictive coding are adopted to significantly increase the coding performance. In AVS3 High Profile, enhanced technologies for intra coding, inter coding, quantization and entropy coding are considered and adopted to further improve rate-distortion (R-D) performance. The overview of the coding tool of AVS3 Main and High Profile is depicted in Table 2, categorized by different coding configurations.

3.1 Two-phase development

AVS3 Main Profile is designed on the top of AVS2 such that the major coding tools in AVS2 are retained and more optimization approaches are additionally designed. More specifically, there are 9 novel coding tools adopted into AVS3 Main Profile under random access (RA) configuration, focusing on intra and inter predictive coding efficiency improvement. Finalized in March 2019, the AVS3 Main Profile achieves over 30% bit-rate reduction over previous video coding standards, AVS2 and HEVC/H.265, for the UHD video coding. It should be noted that the decoding complexity of AVS3 Main Profile is even lower than that of AVS2 and HEVC/H.265, revealing the excellent design philosophy of AVS3. The adoption of such coding tools is based on the comprehensive evaluation of performance, complexity, latency, and hardware design logic. In particular, the tool-off performances of adopted technologies in AVS3 Main Profile are depicted in Table 3, including the efficiency and encoding/decoding complexity. It can be learned that each adopted tool has realized rational trade-off on performance and complexity.

AVS3 High Profile pays more attention to higher compression efficiency when considering the adoption of coding tools and technology development. Starting from April. 2019, over 20 novel coding tools are adopted on the top of AVS3 Main Profile such that more than 10% bit-rate reduction could be obtained. Currently, the AVS3 High Profile has become the industry standard and is advancing into the international standard.

3.2 Intelligent coding tools

The AVS working group has set up an intelligent video coding ad-hoc group (AHG) to investigate the DL solutions for video coding. Currently, this AHG majorly focuses on exploring the coding gain promotion

Table 2 Coding tools and new features adopted in AVS3 standards (the name associated with the abbreviation of each coding tool is depicted)

Configuration	AVS3 Main	AVS3 High
All intra	New partition type, IPF (intra pred filter), TSCPM (two step chroma pred mode), DT_INTRA (intra derived tree)	New partition type, EIPM (extended intra pred mode), SRCC (scan region based coefficient coding), ENHANCE_TSCPM (enhance two step chroma pre mode), IST (implicit selection of transform), MIPF (multiple intra pred filter), PMC (pred from multiple component), MCABAC (multiple context-based adaptive binary coding), EST (enhanced secondary transform), IIP (improvement intra pred), ST_CHROMA (secondary transform for chroma), ESAO (enhanced sample adaptive offset), DBR (deblock refinement), IPF_CHROMA (intra pred filter for chroma), CCSAO (cross-component sample adaptive offset), ALF_SHAPE (adaptive loop filter shape), DEBLOCK_TYPE (deblock type)
Random access	New partition type, IPF (intra pred filter), TSCPM (two step chroma pred mode), DT_INTRA (intra derived tree), AMVR (adaptive motion vector resolution), HMVP (history-based motion vector pred), EMVR (extend motion vector resolution), UMVE (ultimate motion vector expression), AFFINE (affine motion compensation pred), SMVD (symmetric motion vector difference)	New partition type, EIPM (extended intra pred mode), SRCC (scan region based coefficient coding), IST(implicit selection of transform), ENHANCE_TSCPM (enhanced two step chroma pred mode), SBT (subblock-based transform), DMVR (decoder side motion vector refine), BIO (BI-directional Optical flow), INTER-PF (inter pred filter), MVAP (motion vector angular pred), AFFINE_UMVE (affine combined ultimate motion vector expression), MIPF (multiple intra prediction filter), AWP (angular weighted pred), AWP_MVR (angular weighted pred motion vector refine), ETMVP (enhanced temporal motion vector pred), SBTMVP (subblock temporal motion vector pred), DBR (deblock refinement), PMC (prediction from multiple component), MCABAC (multiple context-based adaptive binary coding), EST (enhanced secondary transform), ST_CHROMA (secondary transform for chroma), UMVE_ENHANCE (enhanced ultimate motion vector expression), ESAO (enhanced sample adaptive offset), BGC (Bi-gradient correction), IPF_CHROMA (intra pred filter for chroma), CCSAO (cross-component sample adaptive offset), ASP (affine secondary pred), IIP (improvement intra pred), ALF_SHAPE (adaptive loop filter shape), DEBLOCK_TYPE (deblock type)

Table 3 Tool-off performance evaluation (%) of AVS3 Main Profile adopted coding tools under RA configuration

Coding tool	Y	U	V	EncTime	DecTime
IPF	0.72	1.39	1.19	96	101
TSCPM	0.48	3.89	6.22	100	101
DT_INTRA	0.17	0.70	0.52	96	101
AMVR	1.34	1.71	1.97	77	100
HMVP	0.84	0.95	1.02	85	100
EMVR	0.25	0.25	0.26	90	99
UMVE	1.28	0.45	0.74	95	98
AFFINE	2.18	1.62	1.64	88	101
SMVD	0.12	0.00	0.13	97	101
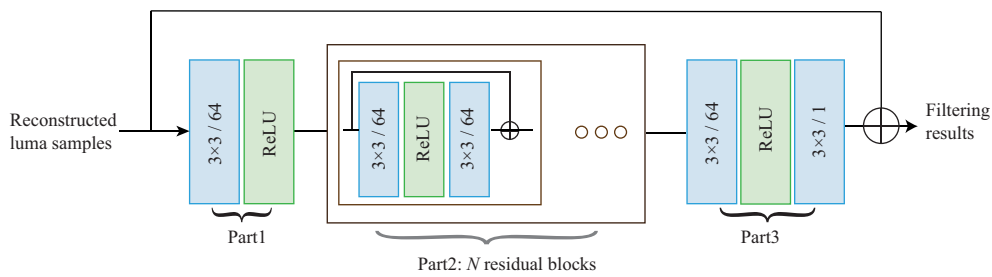
Figure 8 (Color online) The network architecture for luminance in-loop filtering. A global shortcut is appended into the network to make it easier to be trained by capturing the residual. In addition, the residual block is composed of two consecutive convolution layers and a rectified linear unit (ReLU) is the major component of the network. The nested residual learning structure substantially promotes the feature extraction and aggressive ability of the learned filter.
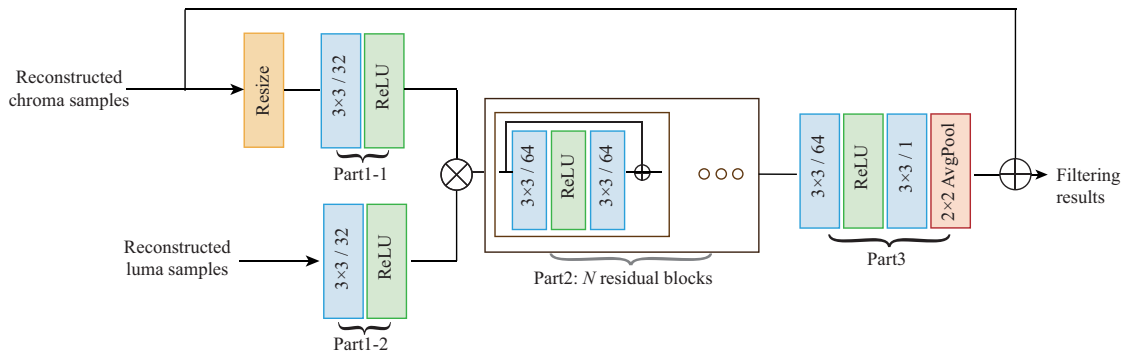
Figure 9 (Color online) The network architecture for chrominance in-loop filtering. In particular, the luma component contains much more textural and visual information, which is utilized as the guidance and further improves the filtering performance for chrominance components.

by substituting hand-crafted modules into deep learning based models in AVS3, e.g., intra prediction, inter prediction, and in-loop filtering. Such methods have great potential to contribute substantial coding gains, bringing the promising opportunity to the development of video coding.

Most of the technical proposals contribute to the involvement of deep learning based in-loop filtering due to its high efficiency and moderate complexity. The compression artifacts can be directly suppressed by the filtering models learned from massive training data, improving the quality of reconstructed frames. More importantly, the enhanced frames could provide high quality temporal reference for the coding of subsequent frames, which is of great benefit to promoting the overall coding efficiency. As shown in Figures 8 and 9, residual based convolutional neural networks for in-loop filtering (CNNLF) are designed for AVS3 standards [42]. Moreover, the filtering network for inter frames is separately trained from video dataset to adapt to the inter coding scenario. The tool-off performance of this technology is around 8% and 10% bit-rate reductions on luma component and chroma components, respectively. All of the trained models of different QPs are used to construct a model candidate list in the deep in-loop filter with adaptive model selection (DAM) [43]. The indication of on/off control as well as the model index is signaled into the bitstream. Such intelligent-based coding tools invoke a novel trend in video coding

research using neural networks, which significantly promotes coding efficiency and extends the horizon of video compression research.

3.3 Technical explorations of AVS3

As video data are advancing towards higher resolution (beyond UHD content) and larger data volume, the AVS working group has officially launched the requirements and explorative research for the display stream compression and conceptual based compression. The former aspect strongly supports the real-time visual systems, with real-time compression, transmission, decompression, and display. The use cases of display stream compression mainly lie in video stream transmission, storage on display devices, broadcasting production, autonomous driving, digital cinema, and medical imaging, etc. The latter one, conceptual based compression, envisions the future video coding as a joint optimization problem of visual understanding and compression, enabling compression domain analysis and semantic preserving based coding techniques.

AVS3 standard targets higher for the third phase to achieve the ever-best R-D performance where the current framework could arrive. Additional coding tools will be adopted on top of Phase-2, especially more adaptive video coding tools and more intelligent neural network based compression methods. Quantization with more adaptive contextual information is under development, and the convolutional neural network (CNN) based in-loop filters and inter prediction approaches have been continuously studied and optimized. Variable resolution reference picture coding is also under investigation to offer temporal and resolution scalability for AVS3.

4 Compression performance

This section presents a comparative study between AVS3 and its predecessor AVS2 standard. In the comparisons, the latest version of AVS2 reference software RD-19.5 is selected for evaluation. Regarding the AVS3 standard, the reference model HPM4.0 (AVS3 Main) and the latest codec HPM-11.0 of AVS3 High are employed. The Bjontegaard delta bitrate (BD-BR) method is adopted for performance evaluation and the peak signal to noise ratio (PSNR) is used as the distortion measure to evaluate the R-D performances.

4.1 R-D performances

Tables 4 and 5 show the coding performance comparison among these standards under different coding configurations. The test conditions follow the common test conditions (CTC) of AVS defined in [44]. Two major coding configurations RA and all intra (AI) are tested. The group of picture (GOP) size in RA structure is set to 16. The fixed quantization parameter (QP) values of 27, 32, 38, and 45 are chosen to encode all test sequences. It can be seen from Tables 4 and 5 that compared with AVS2, over 23.5% and 9.2% bit-rate saving can be achieved for AVS3 Main over the sequences with different resolutions in RA and AI configurations, respectively. For high resolution sequences, the AVS3 Main Profile reaches over 23% improvement beyond the previous AVS standards. It could be observed from Table 4 that the coding performance improvements for UHD contents are larger than 1080P and 720P sequences, because the adopted tools are well designed for UHD contents. The experiments in Table 5 indicate that the AVS3 High Profile provides 33% performance improvement over its predecessors, showing the great potential of the new generation video coding standard. We can also see from the performance comparison that during the development of AVS3, the coding gain in high profile is much larger than that in the main profile. One possible reason for this phenomenon lies in the adoption of cross-component coding tools in AVS3. Guided by the textural information embedded in luma component, significant bit-rate reduction is realized for the chroma components.

AVS standards continuously improve the R-D performances in the past two decades. In addition, it can be easily observed that the latest AVS3 standard can significantly improve coding performance under different bit rates. In particular, the decoder complexity for AVS3 Main and High is well controlled during the development. The light-weighted decoder guarantees AVS3 standard to be a competitive codec towards UHD video coding. The detailed complexity analysis is elaborated in Subsection 4.2.

Moreover, Figure 10 shows the R-D curves for the comparisons using two typical UHD sequences, Tango2 and DaylightRoad2. These figures reflect the R-D behaviors of AVS2, VVC, AVS3 Main, and

Table 4 R-D performance comparison between AVS3 Main Profile (HPM-4.0) and AVS2 (RD-19.5) under AI and RA configuration (unit: %)

Sequence	AI			RA		
	Y	U	V	Y	U	V
UHD	−10.59	−15.39	−15.56	−23.90	−27.54	−29.61
1080P	−8.14	−9.88	−12.81	−23.71	−28.07	−29.06
720P	−8.92	−8.30	−8.83	−22.96	−27.23	−26.47
Overall	−9.22	−11.19	−12.40	−23.52	−27.62	−28.38
Encoding time	1331			392		
Decoding time	100			66		

Table 5 R-D performance comparison between AVS3 High Profile (HPM-11.0) and AVS2 (RD-19.5) under AI and RA configuration (unit: %)

Sequence	AI			RA		
	Y	U	V	Y	U	V
UHD	−19.15	−31.00	−30.34	−33.97	−43.61	−46.96
1080P	−17.66	−30.37	−29.32	−34.10	−47.69	−44.93
720P	−17.78	−26.36	−24.20	−35.25	−48.36	−45.37
Overall	−18.19	−29.24	−27.95	−34.44	−46.55	−45.75
Encoding time	2824			561		
Decoding time	102			81		
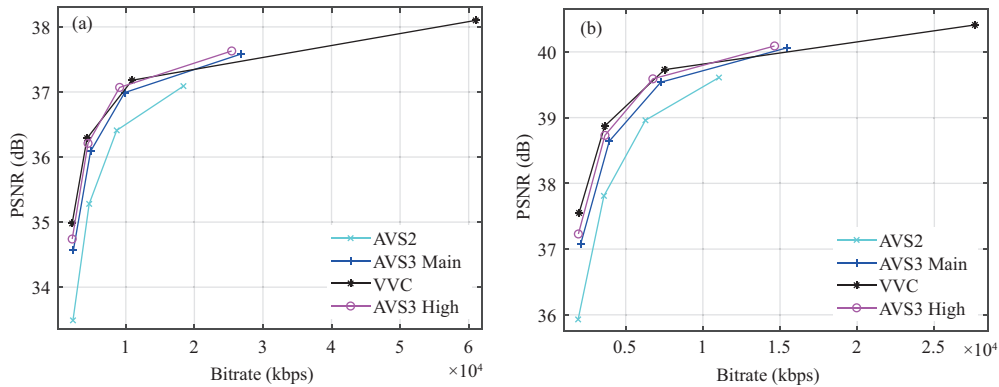
Figure 10 (Color online) R-D curves of AVS3 Main (HPM-4.0), VVC (VTM-10.0), AVS3 High (HPM-11.0), and AVS2 for UHD video coding under RA configuration using test video sequences (a) DaylightRoad2 and (b) Tango2.

AVS3 High Profile under RA configuration. It can be easily found that significant coding performances are obtained from AVS2 to AVS3 standard. For both of the low bit-rate and high bit-rate coding scenarios, the AVS standards realize consistent and significant improvement during the evolution and development. In addition, the more coding gain is obtained under low bit-rates when comparing the R-D curves of the two different profiles of AVS3.

To provide comprehensive interpretation of the technical comparison between the AVS3 standard and the latest H.266/VVC standard, we further show the algorithmic descriptions of the coding tools associated with their tool-off performances in Tables 6 and 7.

As aforementioned, AVS3 has adopted deep network based video coding tools as a pioneer investigation for the intelligent coding. In particular, an artificial intelligence branch is established on the top of the reference model (Mod AI). We also provide the simulation results of the Mod AI against the previous AVS standard in Table 8. The coding performances under RA configuration are evaluated and reported. On average, the intelligent tool based AVS3 model achieves over 40% bit-rate reduction than the previous generation AVS standard for both of the luma and chroma components. It is also worth noting that the coding performance improvements are balanced between different color components. It could be also observed from Table 8 that the coding gain of different video sequences shows consistent R-D behavior, which indicates that the intelligent coding tool generalizes well on diverse video contents and could be applied to ubiquitous coding scenarios.

**Table 6** AVS3 reference software HPM10.0 coding tool performances

Coding module	Tool name	Abbreviation	Performance (BD-Rate) (%)		
			Y	U	V
Coding unit partition	Quadtree with binary tree and extended Quadtree	QT+BT+EQT	−	−	−
Intra prediction	Intra prediction filter	IPF	0.72	1.39	1.19
	Two step cross-component prediction mode	TSCPM	0.48	3.89	6.22
	Derived tree	DT	0.17	0.70	0.52
	Extended intra prediction modes	EIPM	1.29	1.21	1.27
	Improved intra prediction	IIP	0.10	0.23	0.23
	Prediction from multiple cross-components	PMC	0.10	1.45	−0.50
Inter prediction	History based motion vector prediction	HMVP	0.84	0.95	1.02
	Ultimate motion vector expression	UMVE	1.28	0.45	0.74
	Adaptive motion vector resolution	AMVR	1.34	1.71	1.97
	Extended motion vector resolution	EMVR	0.25	0.25	0.26
	Symmetric motion vector difference	SMVD	0.12	0.00	0.13
	Affine motion compensation	AMC	2.18	1.62	1.64
	Affine secondary prediction	ASP	0.58	0.27	0.27
	Angular weighted prediction	AWP	0.88	1.22	1.40
	Subblock temporal motion vector prediction	SBTMVP	0.10	0.17	0.17
	Enhanced temporal motion vector prediction	ETMVP	0.08	−0.07	−0.04
	Motion vector angular prediction	MVAP	0.16	0.23	0.21
	Decoder-side motion vector refinement	DMVR	0.55	0.60	0.72
	BI-directional optical flow	BIO	1.33	0.68	0.47
	Inter prediction filter	INTERPF	0.57	0.34	0.48
	Bi-directional gradient correction	BGC	0.57	0.17	0.48
Transform/quantization	Enhanced secondary transform	EST	0.21	0.12	0.14
	Subblock transform	SBT	0.43	−0.14	−0.13
	Implicit selection of transforms	IST	0.92	0.78	0.83
Entropy coding	Scan region based coefficient coding	SRCC	2.80	1.35	1.34
	Multiple hypothesis model based entropy coding	MEC	0.25	0.72	1.39
In-loop filter	Enhanced sample adaptive offset	ESAO	0.54	0.20	0.11
	Cross-component sample adaptive offset	CCSAO	−0.20	9.11	7.09
	Enhanced deblocking filter	EDBF	0.12	0.28	0.00
	Enhanced adaptive loop filter	EALF	0.54	−0.27	−0.03

Table 7 H.266/VVC reference software VTM10.0 coding tool performances

Coding module	Tool name	Abbreviation	Performance (BD-Rate) (%)		
			Y	U	V
Coding unit partition	Multi-type tree	MTT	–	–	–
Intra prediction	Chroma separate tree	CST	0.12	4.12	4.57
	Cross-component linear model	CCLM	1.02	11.52	12.59
	Multi-reference line prediction	MRLP	0.17	0.08	0.10
	Matrix based intra prediction	MIP	0.33	0.35	0.37
	Intra sub-partitioning	ISP	0.28	0.29	0.33
Inter prediction	Subblock-based temporal merging candidates	SBTMC	0.43	0.29	0.38
	Geometry partition mode	GPM	0.67	1.14	1.16
	Merge with motion vector difference	MMVD	0.52	0.44	0.49
	Adaptive motion vector resolution	AMVR	1.42	2.15	2.36
	Combined intra/inter prediction	CIIP	0.26	0.00	−0.02
	Symmetric motion vector difference	SMVD	0.25	0.23	0.23
	Bi-prediction with CU weights	BCW	0.40	0.46	0.46
	Affine motion compensation	AMC	3.10	2.30	2.15
	Prediction refinement using optical flow	PROF	0.48	0.14	0.08
	Decoder motion vector refinement	DMVR	0.83	1.11	1.14
	Bi-directional optical flow	BDOF	0.76	0.33	0.26
Transform/quantization	Multiple transform set	MTS	0.75	0.66	0.64
	Low frequency non-separable transform	LFNST	0.70	0.78	1.08
	Subblock transform	SBT	0.41	−0.03	−0.02
Entropy coding	Dependent quantization	DQ	1.60	0.95	0.66
	Joint coding of chrominance residuals	JCCR	0.53	0.40	0.14
In-loop filter	Sampled adaptive offset	SAO	0.08	0.14	0.31
	Cross-component adaptive loop filter	CCALF	-0.13	13.88	13.73
	Adaptive loop filter	ALF	4.34	19.31	19.06
	Luma mapping with chroma scaling	LMCS	1.38	1.14	0.99

Table 8 R-D performance comparing AVS3 artificial intelligence branch with AVS2 (RD-19.5) and with VVC (VTM-10.0) under RA configuration (unit: %)

Sequence	vs. AVS2			vs. VVC		
	Y	U	V	Y	U	V
UHD	−39.90	−44.40	−45.84	−2.04	−6.78	0.32
1080P	−40.34	−47.41	−46.62	−3.73	0.63	−4.30
720P	−41.58	−50.15	−48.09	−2.55	5.25	−2.17
Overall	−40.60	−47.32	−46.85	−2.77	−0.30	−2.05
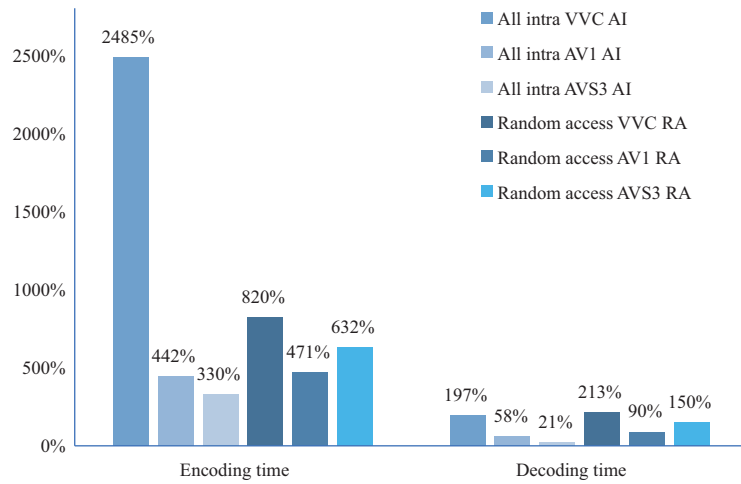
Figure 11 (Color online) The encoding and decoding complexity increasing of VVC [46], AV1 [47] and AVS3 [3] codecs (AVS3 Main Profile) with respect to HEVC [45] standard. Comparisons are based on the reference software of these standards. Regarding encoder, significant encoding time increasing is observed for all of the emerging standards among different coding structures. In addition, VVC has the highest encoding and decoding complexity against AVS3 and AV1 standards.

4.2 Complexity analysis

With the upgrade of the coding tools and the advancement of the optimization algorithms, a series of new features are supplemented into AVS3 standard, such as the patch-based encoding structure and the size expansion of the CUs. Furthermore, the reference frame structure and relationship in AVS3 become more complicated. All of the factors result in higher computational complexity and resources consumption. To quantitatively evaluate the complexity and performance trade-off, the run-time comparison for reference software among HEVC [45], VVC [46], AV1 [47], and AVS3 [3] Main Profile is compared and shown in Figure 11. The common test condition sequences of VVC are adopted for comparison.

In general, the soaring computational complexity among the new generation of video coding standards has become a severe challenge for the industrialization process. Consistent complexity increasing could be observed in different prediction structures, which indicates that the tools adopted are selected in both intra and inter coding scenarios. It is additionally observed that VVC has the highest complexity in terms of both encoding and decoding time compared to AV1 and AVS3. The all intra coding of VVC takes 24 times than the HEVC all intra coding, while the number for random access is over 8. This is because a much more number of coding tools and complex algorithmic logic are utilized. For AI configuration, AVS3 achieves encouraging complexity control against other codecs and it is much faster than the HEVC decoder. Regarding the RA configuration, AVS3 enjoys better complexity control and trade-off compared to the VVC standard.

5 Real-world applications: towards 8K UHD video coding

Nowadays, the 4K and 8K UHD ecosystems are being built rapidly including UHD content creation, UHD TV broadcasting, and UHD on-demand video streaming. Therefore, the requirements and applications of the UHD video compression are growing exponentially. With the emerging AVS3 standard, impacts have been made in the context of 8K UHD video coding and related video applications. Targeting at higher efficiency UHD video coding with low energy cost and transparent patent policy, AVS3 has attracted

tremendous attention from industry ever since the finalization of AVS3 Main Profile in March, 2019. The designation and performance-complexity trade-off naturally fulfill the demands and requirements of different use cases. In this section, several representative applications of AVS3 standard are illustrated to reflect its superiority, adaptivity, and scalability. In particular, we describe the wide deployment of AVS3 from industrial-grade 8K real-time software codecs to 8K decoder chip for smart-phone, from 8K UHD TV broadcasting service to worldwide large-scale events support.

5.1 AVS3 8K UHD real-time codecs

8K UHD real-time AVS3 encoder. Based on the Intel scalable video technology (SVT) architecture, a novel AVS3 8K real-time software encoder has been presented, namely SVT-AVS3 [48]. The SVT-AVS3 encoder achieves an optimal trade-off among coding performance, encoding speed, and perceptual visual quality. Through different level systematic designs, SVT-AVS3 is supposed to maximize the utilization of various high-performance machines. In particular, SVT-AVS3 carefully considers multidimensional parallelism and perceptual based rate-distortion optimization. Parallelism in SVT-AVS3 allows encoder to scale its efficiency-complexity performance properly in response to the computation and memory constraints while maintaining a scalable degradation in video quality, enabled by single-instruction-multiple-data (SIMD) optimization.

• Process-level parallelism. In SVT-AVS3, the complex encoding tasks are divided and distributed into several computing cores. Each coding core is instantiated into multiple threads of process. The data processing-oriented threads such as analysis are mainly designed to process input data. The control-oriented threads such as the picture manager process are mainly designed to synchronize the operational tasks of encoder. The encoding operations are balanced into different encoding threads.

• Instance-level parallelism. SVT-AVS3 allows an encoding application to run up to 6 independent instances with different configurations. The computational resources are still handled by each instance but are more friendly for system scheduling.

• Segment-level parallelism. The video segments are supposed to be processed by separate processor cores which further improve the utilization of computational resources. Through input video splitting and stream merging, SVT-AVS3 can distribute the input stream to several independent CPUs and output into a single stream.

• Picture-level parallelism. Based on standard-compliant GOP configuration, SVT-AVS3 adopts a more flexible frame management scheme. Picture description and data pointers are completely handled by a picture control set (PCS). Reference information is separated into another data structure and stored into a reference list (RL).

SVT-AVS3 can support 8K UHD real-time encoding with 10-bit/sample using around 100 Mbps bandwidth, which provides multiple quality vs. coding speed presets from M0 to M11 to adapt to variable coding scenarios, in which M0 encapsulates all features in AVS3 with highest compression efficiency while M11 realizes 8K real-time encoding. SVT-AVS3 divides its coding process into two different loops based on the complexity and logic analysis of AVS3, the open loop and the closed loop, as shown in Figure 12. Before the formal encoding process, a pre-analysis procedure is equipped to determine the high-level computation resource scheduling and allocation mechanism. Together with the two kinds of coding loops, SVT-AVS3 could balance most of its own computing resources to attain a better scheduling strategy than the conventional CPU strategies. In addition, almost all optional coding tools and algorithmic procedure in SVT-AVS3 are deeply decoupled and parameterized. These function sets provide a robust mechanism for converting diverse application scenarios. In addition, SVT-AVS3 also authorizes few adjustable overhead for structured features. That information visibly enhances the subjective quality particularly under low bit rate conditions. Regarding the encoding quality and speed, SVT-AVS3 achieves 22.9% bit-rate reduction using PSNR as distortion metric while 29.8% using SSIM, when compared with x265 on the best quality mode. SVT-AVS3 is about 5 times faster than x265 ultrafast preset level on Intel Xeon Platinum 8180 platform. For 8K UHD 10-bit sequences, SVT-AVS3 could encode more than 40 frames per second with a single 8180 processor. With dual 8180 processors and a GOP-level parallelism outside encoder, SVT-AVS3 could provide an encoding speed of 75 frames per second.

8K UHD real-time AVS3 decoder. Multiple investigations have been made for real-time AVS3 decoders using optimized processing and design principle, advanced data structure, SIMD instructions, multi-threading techniques, and heterogeneous computing frameworks [49, 50]. To increase GPU utilization, high parallel and low memory access latency schemes are carefully designed for each module in
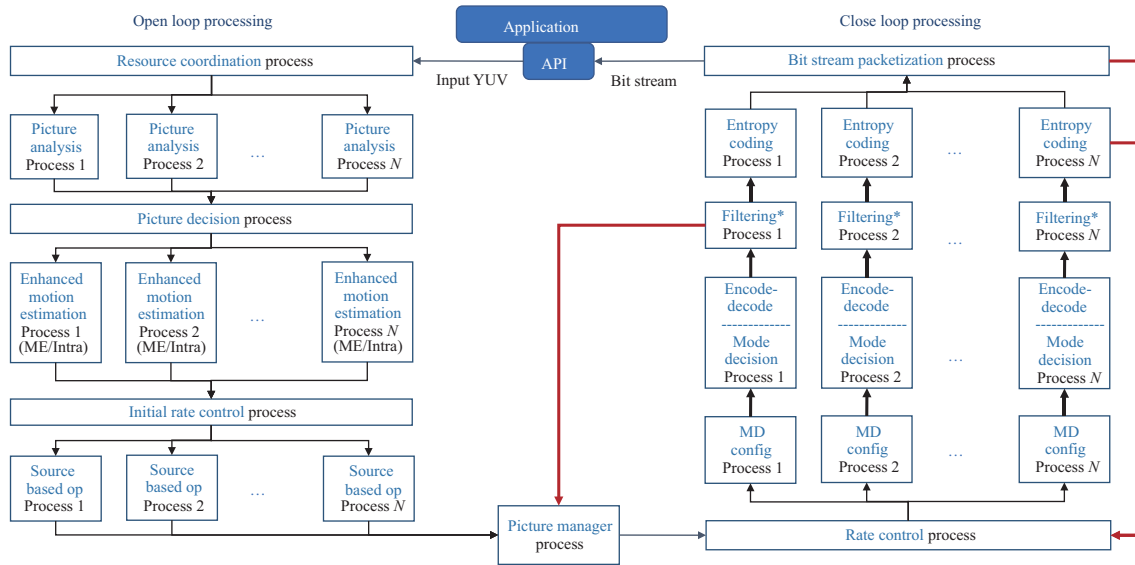
Figure 12 (Color online) The parallel coding architecture of SVT-AVS3. The input videos are first analyzed by the open loop processing module and then encoded by the closed loop processing module.

AVS3. The optimized AVS3 decoder archives 113 fps for 4K bitstreams decoding with a single NVIDIA GeForce RTX 2080Ti GPU. When the bit-rate reaches 300 Mbps, the decoding speed is still higher than 50 fps. Regarding the 8K bitstream decoding, an average frame rate of 47 fps is obtained, which is 78 times faster than the decoder of AVS3 reference software. In addition, evidence has been shown that optimized software decoder could be deployed on ARM platform for real-time mobile utilities [50].

5.2 8K decoder SoC

In addition to software based applications, the AVS3 standard has been practically applied to hardware support for consumer electronics such as set-top box or smart phones. In September 2019, Hisilicon Inc. has officially released a new end-to-end solution for 8K contents based on the world's first AVS3 8K decoder chip Hi3796CV300 in International Broadcasting Convention (IBC), which greatly promotes the 8K applications of AVS3. The Hi3796CV300 is the world's first user-end product to support 8K@120 fps AVS3 video decoding. This chip also supports HEVC and HiSilicon claims that this is the industry's first 8K@120 fps ultra-HD video and voice platform, with HEVC and AVS3 video decoding. This chip is powered by an octa-core CPU with A73 cores and a Mali G52 NP6 GPU. This chip also has commendable neural processing units (NPUs) delivering up to 4 TOPs for AI and machine learning workloads. Such technology lays the solid foundations for modern smart home and other internet-of-things applications.

5.3 8K UHD TV broadcasting

The UHD TV broadcasting has been realized using AVS standard. For 4K UHD TV broadcasting, the China Central Television (CCTV), the state television of China, has launched the first 4K UHD live channel using AVS2 standard in 2018. Towards 8K UHD, the signals are tested and transmitted from the headquarters of China Media Group (CMG) in Beijing to display terminals at the venue of the Mobile World Congress 2019 in Shanghai. In January 2021, CCTV has successfully realized 8K UHD TV broadcasting channel. As depicted in Figure 11, the Spring Festival Gala 8K live show has been provided over 10 different cities in China in February, 2021, enabled by the ultra-fast transmission speed of 5G networks. The 8K UHD channel has brought brand new user experience to the cable network users all over China.

5.4 Patent policy

The state-of-the-art AVS3 standard opens a new era for 8K UHD video coding, especially for its multi-dimensional superiority in coding efficiency, deployment flexibility, and patent policy. The IPR policy of AVS is consistent and transparent. AVS3 is also fully compliant with the IPR policy of IEEE standards

and AVS working group. The licensing terms are also considered during the adoption of proposals for AVS standards; in particular equal technical importance is assigned to all the techniques.

AVS IPR Policy was developed in 2004 by an international team of patent licensing experts and patent attorneys, of which the representatives come from semiconductor, consumer, communication, and computer industries. AVS workgroup establishes a low risk innovation model which allows to build up the licensing model before standardization. The advantages of AVS simple licensing model lie in the following aspects:

- One simple patent pool;
- One-stop licensing model;
- 1 Chinese Yuan per device for hardware decoder (AVS1/AVS+/AVS2 Patent Pool);
- Annual royalty cap; fixed annual fee option (AVS1/AVS+/AVS2 Patent Pool);
- Royalty free for software decoder in Internet applications (AVS1/AVS+/AVS2 Patent Pool);
- Royalty free for content encoding (AVS1/AVS+/AVS2 Patent Pool).

In this innovation working model, all proponents who would like to propose technologies to AVS, shall disclose their patents and declare their licensing commitment ("rand royalty-fee" or "agree to join in the patent pool"). AVS standards are explicitly designed to use the technologies from AVS member only, or use royalty free technologies.

6   Conclusion

This paper provides a comprehensive overview of the development of AVS video coding standards and their real-world applications with an emphasis on the technical contributions over the past two decades. Plenty of high efficiency and light-weight coding tools have been investigated and adopted for diverse requirements. It has also been recognized by both industry and academia that the AVS video coding standards, especially the emerging AVS3 standard, reach a level of maturity and become one of the leading technologies in video industry. AVS3 is also opening a new era for 8K UHD video coding in which end-to-end solutions and commercial-level AVS3 codecs have been incorporated and established. Several directions are currently under investigation for future research and applications, including neural-network based video coding technology and machine-vision oriented video coding. Related standardization work has been initiated by the AVS workgroup.

References

1 Fan L, Ma S, Wu F. Overview of AVS video standard. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), 2004. 423–426

2 Ma S, Huang T, Reader C, et al. AVS2—making video coding smarter [standards in a nutshell]. IEEE Signal Process Mag, 2015, 32: 172–183

3 Zhang J, Jia C, Lei M, et al. Recent development of AVS video coding standard: AVS3. In: Proceedings of IEEE Picture Coding Symposium (PCS), 2019. 1–5

4 Wang M, Li J, Zhang L, et al. Extended coding unit partitioning for future video coding. IEEE Trans Image Process, 2019, 29: 2931–2946

5 Wang L, Niu B, Wei Z, et al. CE1: technology of block. In: Proceedings of the 67th Audio and Video Coding Standard Meeting, AVS_M4540, Shenzhen, 2018

6 Shi L, Lou J, Yu L. The new intra prediction method for 8×8 block. In: Proceedings of the 6th AVS Meeting, AVS_M1152, Hangzhou, 2003

7 Wang Q, Zhao D, Gao W. A new intra prediction. In: Proceedings of the 6th AVS Meeting, AVS_M1161, Hangzhou, 2003

8 Zhang N, Yin B, Kong D, et al. Spatial prediction based intra-coding. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), 2004. 97–100

9 Yu Q, Zhao L, Ma S. Multiple directional intra prediction modes for intra coding. In: Proceedings of the 43rd AVS Meeting, AVS_M3001, Beijing, 2012

10 Cai Y, Yu Q, Wang S, et al. A bilinear mode for intra prediction. In: Proceedings of the 43rd AVS Meeting, AVS_M2999, Beijing, 2012

11 Piao Y, Lee S, Kim C. Adaptive MPM for intra mode coding. In: Proceedings of the 47th AVS Meeting, AVS_M3234, Shenzhen, 2013

12 Lei M, Luo F, Wang S, et al. CE2-related: extended intra prediction modes. In: Proceedings of the 70th AVS Meeting, AVS_M4993, Haikou, 2019

13 Wang F, Xie Z, Yuan Q, et al. CE1-related: the optimization of spatial angular weighted prediction. In: Proceedings of the 75th AVS Meeting, AVS_M6099, 2021

14 Piao Y, Chen J, Lee S, et al. Intra coding of AVS2 video coding standard. In: Proceedings of IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2014. 1–6

15 Wang Y, Xu X, Liu S, et al. CE1-related: multiple reference samples filters for intra prediction. In: Proceedings of the 71st AVS Meeting, AVS_M5079, Shenzhen, 2019

16 Fan K, Wang R, Li G, et al. Efficient prediction methods with enhanced spatial-temporal correlation for HEVC. IEEE Trans Circ Syst Video Technol, 2018, 29: 3716–3728

17 Li J, Wang M, Zhang L, et al. Sub-sampled cross-component prediction for emerging video coding standards. IEEE Trans Image Process, 2021, 30: 7305–7316

18 Li J, Zhang L, Zhang K, et al. Prediction with multi-cross component. In: Proceedings of International Conference on Multimedia & Expo Workshops (ICMEW), 2020. 1–6

19 Yue L, Yu L. F framece: a forward double-hypothesis prediction coding scheme. In: Proceedings of the 48th AVS Meeting, AVS_M3326, Beijing, 2014

20 Ji X, Zhao D, Gao W, et al. New bi-prediction techniques for B pictures coding [video coding]. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), 2014. 101–104

21 Ling Y, Zhu X, Yu L. Multi-hypothesis mode for AVS2. In: Proceedings of the 47th AVS Meeting, AVS_M3271, Shenzhen, 2013

22 Kim I K, Lee S, Piao Y, et al. Directional multi-hypothesis prediction (DMH) for AVS2. In: Proceedings of the 45th AVS Meeting, AVS_M3094, Taicang, 2013

23 Fang S, Sun Y, Chen F, et al. Ce3.3: motion vector angle prediction. In: Proceedings of the 70th AVS Meeting, AVS_M4926, Haikou, 2019

24 Li J, Wang M, Zhang L, et al. History-based motion vector prediction for future video coding. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), 2019. 67–72

25 Ma J, Ma S, An J, et al. Progressive motion vector resolution. In: Proceedings of the 44th AVS Meeting, AVS_M3049, Luoyang, 2013

26 Sun Y, Chen F, Wang L, et al. Angular weighted prediction for next-generation video coding standard. In: Proceedings of IEEE International Conference on Multimedia and Expo (ICME), 2021. 1–6

27 Xiu X, Kuo C W, Chen W, et al. Overlapped block motion compensation for AVS3. In: Proceedings of the 76th AVS Meeting, AVS_M6105, 2021

28 Xu W, Zhao Y, Yang H. Inter prediction filter. In: Proceedings of the 69th AVS Meeting, AVS_M4812, Chengdu, 2019

29 Lu X, Yang H. Affine motion compensation in AVS3. In: Proceedings of the 66th AVS Meeting, AVS_M4451, Changchun, 2019

30 Xu W, Yang H. Simplified decoder side motion vector refinement. In: Proceedings of the 69th AVS Meeting, AVS_M4813, Chengdu, 2019

31 Wang F, Ouyang X, Lu Z, et al. Bi-directional optical flow. In: Proceedings of the 69th AVS Meeting, AVS_M4762, Chengdu, 2019

32 Mao X, Wang Y, He Y. Adaptive block size coding for AVS-X profile. In: Proceedings of the 25th AVS Meeting, AVS_M2372, Xiamen, 2008

33 Wang Z, Wang R, Fan K, et al. The weighted quantization for AVS3. In: Proceedings of the 67th AVS Meeting, AVS_M4667, Shenzhen, 2018

34 Zheng J, Zheng X, Meng X. Parameter-weighted quantization in AVS X-profile. In: Proceedings of the 18th AVS Meeting, AVS_M1878, Beijing, 2006

35 Wang Q, Zhao D B, Gao W. Context-based 2D-VLC entropy coder in AVS video coding standard. J Comput Sci Technol, 2006, 21: 315–322

36 Zhang L, Wang Q, Zhang N, et al. Context-based entropy coding in AVS video coding standard. Signal Processing-Image Commun, 2009, 24: 263–276

37 Wang J, Wang X, Ji T, et al. Transform coefficient coding design for AVS2 video coding standard. In: Proceedings of 2013 Visual Communications and Image Processing (VCIP), 2013. 1–6

38 Lv Z, Piao Y, Wu Y, et al. Scan region-based coefficient coding in AVS3. In: Proceedings of 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2020. 1–5

39 Jian Y, Zhang J, Luo F, et al. CE6-1: one enhanced SAO filtering algorithm. In: Proceedings of the 73rd AVS Meeting, AVS_M5373, Teleconference, 2020

40 Kuo C W, Xiu X, Chen W, et al. CE4-1: cross-component sample adaptive offset. In: Proceedings of the 74th AVS Meeting, AVS_M5800, Teleconference, 2020

41 Pan D, Sun Y, Chen F, et al. Optimization on ALF filter shape. In: Proceedings of the 74th AVS Meeting, AVS_M5589, Teleconference, 2020

42 Lin K, Jia C, Zhao Z, et al. Residual in residual based convolutional neural network in-loop filter for AVS3. In: Proceedings of IEEE Picture Coding Symposium (PCS), 2019. 1–5

43 Li J, Wang M, Li Y, et al. Deep in-loop filter with adaptive model selection for AVS. In: Proceedings of the 76th AVS Meeting, AVS_M6356, Teleconference, 2021

44 Fan K. AVS3-p2 common test condition. In: Proceedings of the 68th AVS Meeting, AVS_N2654, Qingdao, 2019

45 Sullivan G J, Ohm J R, Han W J, et al. Overview of the high efficiency video coding (HEVC) standard. IEEE Trans Circ Syst Video Technol, 2012, 22: 1649–1668

46 Bross B, Wang Y K, Ye Y, et al. Overview of the versatile video coding (VVC) standard and its applications. IEEE Trans Circ Syst Video Technol, 2021, 31: 3736–3764

47 Han J, Li B, Mukherjee D, et al. A technical overview of AV1. Proc IEEE, 2021, 109: 1435–1462

48 Ren H, Jia C, Luo F, et al. SVT-AVS3: scalable video technology of AVS3. In: Proceedings of IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2020. 267–270

49 Han X, Jiang B, Wang S, et al. GPU based real-time UHD intra decoding for AVS3. In: Proceedings of International Conference on Multimedia & Expo Workshops (ICMEW), 2020. 1–6

50 Wang Z, Han B, Wang R, et al. UAVS3D: fast decoder for the 3rd generation audio video coding standard (AVS3). In: Proceedings of 4th International Conference on Digital Signal Processing (ICDSP), 2020. 51–55