DOI: 10.3969/j. issn. 1673-5188. 2016. 01. 004
http://www.cnki.net/kcms/detail/34.1294.TN.20160122.1848.004.html, published online January 22, 2016

SHVC, the Scalable Extensions of HEVC, and Its Applications

Yan Ye[1], Yong He[1], Ye-Kui Wang[2], and Hendry[2]
(1. InterDigital Communications, Inc., San Diego, CA 92121, USA;
2. Qualcomm Incorporated, San Diego, CA 92121, USA)

Abstract

This paper discusses SHVC, the scalable extension of the High Efficiency Video Coding (HEVC) standard, and its applications in broadcasting and wireless broadband multimedia services. SHVC was published as part of the second version of the HEVC specification in 2014. Since its publication, SHVC has been evaluated by application standards development organizations (SDOs) for its potential benefits in video applications, such as terrestrial and mobile broadcasting in ATSC 3.0, as well as a variety of 3GPP multimedia services, including multi-party multi-stream video conferencing (MMVC), multimedia broadcast/multicast service (MBMS), and dynamic adaptive streaming over HTTP (DASH). This paper provides a brief overview of SHVC and the performance and complexity analyses of using SHVC in these video applications.

Keywords

HEVC; SHVC; broadcasting; video conferencing; video streaming

1 Introduction

High-Efficiency Video Coding (HEVC) [1] is the state-of-the-art video coding standard developed by the Joint Collaborative Team on Video Coding (JCT-VC) of ISO/IEC JTC 1 SC 29/WG 11 MPEG and ITU-T Q6/16 VCEG. Finalized in January 2013, the first version of HEVC achieved more than 50% bit rate reduction over its predecessor H.264/MPEG-4 part 10 Advanced Video Coding (H.264/AVC) [2] at comparable subjective quality [3]. An overview of HEVC can be found in [4].

The first version of HEVC provides support for temporal scalability. To support other types of scalabilities, such as spatial scalability and quality scalability, the ISO/IEC MPEG and ITU-T VCEG issued a joint call for proposals [5] for scalable video coding extensions of HEVC (SHVC) in July 2012. In October 2012, twenty responses were received from companies, research institutes, and universities worldwide, and the development of the SHVC standard officially started. In July 2014, SHVC was finalized as part of the second version of HEVC [6], [7], which also includes the multiview extensions of HEVC (MV-HEVC) and the range format extensions of HEVC (RExt). An SHVC test model document describing the non-normative aspects of SHVC, including encoder description, as well as the reference software continued to evolve after the normative SHVC specification was finalized. The latest SHVC test model (SHM 10) document and reference software can be found in [8] and [9], respectively. The common conditions under which the performance of SHVC is tested can be found in [10].

In recent years, video entertainment habits have changed significantly. Smartphones, tablets, and other portable devices are equipped with increasingly more powerful computing capabilities and faster network connections. These devices provide rich platforms for video and multimedia applications. Instead of sitting in front of the TV and watching pre-scheduled programs provided by free-to-air or cable networks, people are spending more time consuming video content on-demand through a wide variety of devices, such as living room TVs, smartphones, tablets, and laptops. The N-screen scenario, where video content is generated from and distributed to different terminals with a wide range of capabilities, has become common. Furthermore, more collaboration and communication in the workplace and at home involves video chat, multi-party video conferencing, and telepresence. In light of the significant increase in device and network heterogeneity, scalable video coding can potentially make networks more efficient and resilient to errors. For this reason, since SHVC was finalized in 2014, various application standards development organizations (SDOs) have quickly taken up the tasks of evaluating the potential benefits of supporting SHVC in their applications.

The Advanced Television Standardization Committee (ATSC) was established in the early 1980s. The most widely used standard developed by ATSC is ATSC1.0, which is used for digital television transmission in the United States, Canada, Mexico, South Korea, and a few other North and South American countries. Since 2013, the committee has been developing the ATSC 3.0 standard, with the goal of providing more services to the viewer with increased bandwidth efficiency and better compression. Because broadcasters need to transmit video programs in a variety of formats, including standard definition (SD) [11], high definition (HD) [12], and ultra-high definition (UHD) [13], scalable video coding can provide better coding efficiency compared to transmitting these various video formats independently using simulcast. After careful review of the coding performance and complexity of SHVC, the committee recently decided to adopt the support of SHVC into ATSC 3.0. Commercial deployment of ATSC 3.0 is expected to emerge within the next few years.

The 3GPP is a collaboration between groups of telecommunications associations. 3GPP has developed a number of mobile communications standards that are widely deployed around the globe, including GSM, Universal Mobile Telecommunications System (UMTS), High Speed Packet Access (HSPA), and most recently, 4G Long Term Evolution (LTE). 3GPP SA WG4 Codec (SA4) specifies speech, audio, video, and multimedia codecs in both circuit-switched and packet-switched environments. As mobile and portable devices become main consumption platforms for video and multimedia applications, much pressure is put on wireless network operators to provide rich multimedia experience to a wide range of devices with maximum bandwidth efficiency. Scalable video coding can increase the ability of service providers to adapt to the capabilities of customer devices and fluctuating network conditions. Scalable video coding can also provide better error resilience because it combines naturally with unequal error protection mechanisms to better combat error-prone wireless channels. For this reason, 3GPP SA4 established a video-enhancements study item with a focus on the performance and complexity of SHVC in a number of mobile video applications, including multiparty multistream video conferencing (MMVC), multimedia broadcast/multicast service (MBMS), and 3GPP dynamic adaptive streaming over HTTP (3GP-DASH).

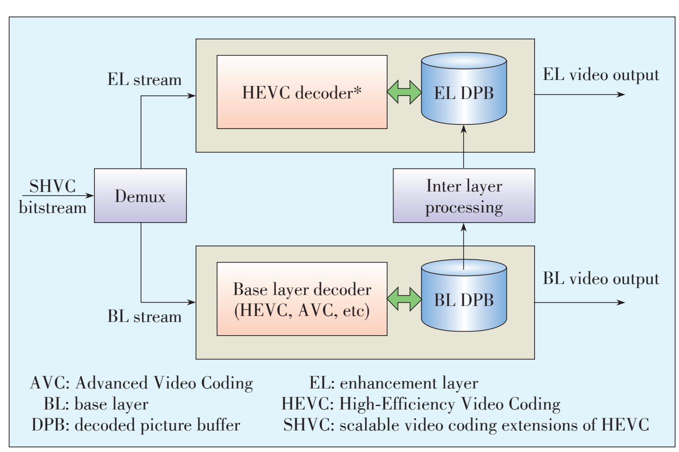
The remainder of this paper is organized as follows. In section 2, SHVC architecture is briefly reviewed. In section 3, the performance of SHVC for terrestrial and mobile broadcasting in ATSC 3.0 is discussed. In section 4, the performance of SHVC for a number of 3GPP video applications is discussed. Section 5 concludes the paper.

2 SHVC

In scalable video coding, interlayer prediction (ILP) is a powerful tool for improving the coding efficiency of enhancement layers (ELs). ILP involves predicting an EL picture using a base layer (BL) or another lower reference layer picture. Take a two-layer scalable coding system that consists of one BL and one EL for example. SHVC uses the so-called "reference index" framework for efficient ILP. In the reference index framework, the reconstructed picture of the BL is treated as an interlayer reference picture (ILRP). The existing reference index signaling that is already part of the single-layer HEVC codec is used to identify whether the block-level prediction comes from the BL or current EL. Such an ILP method is similar in principle to the multiview extension of H.264/AVC (Annex H in [2], also commonly referred to as MVC) and MV-HEVC. This reference-index-based framework of SHVC is fundamentally different from its predecessor, the scalable extension of H.264/AVC (Annex G in [2], also commonly referred to as SVC), which instead relies on a block-level flag to indicate whether an EL block is predicted from the BL or current EL.

Fig. 1 shows the SHVC codec architecture from the decoder's perspective using a two-layer system as an example. The BL reconstruction is retrieved from the BL decoded picture buffer (BL DPB). If necessary, appropriate interlayer processing is done to the reconstructed BL picture to obtain the interlayer reference picture. The ILRP is put into the EL DPB as a long-term reference picture and is used with the EL temporal reference pictures for EL coding.

There are a number of design benefits with the reference index framework. First, all block-level logic of the EL codec is kept the same as that of a single-layer HEVC codec. Changes made to support the EL codec are limited to the slice header level and above; in other words, they are limited to the high-level syntax (HLS). Therefore, the EL decoder is labelled an HEVC decoder* (Fig. 1). Making HLS-only changes enables the existing ASIC design of an HEVC codec to be reused to the greatest possible extent to implement an SHVC codec. Second, the BL codec in Fig. 1 can operate as a black box because the scalable coding of the EL only requires the reconstructed BL pictures. This allows earlier-generation codecs, such as H.264/AVC, to be used in the BL for backward compatibility. The more efficient HEVC codec is used in the EL to improve cod-



▲**Figure 1. SHVC decoder architecture with two layers. The EL decoder has the same block-level logic as a single-layer HEVC decoder.**

ing performance. Finally, the scalable system in Fig. 1 is compatible with MV-HEVC. Although SHVC and MV-HEVC started out as different efforts, a unified architecture of the two extensions is desirable. Once one of these two has been implemented, the other can be easily added, and this can increase the chances that both extensions will be commercially used.

In terms of computation complexity, the architecture in Fig. 1 is based on the multi-loop decoding design. This means that all lower reference layers need to be fully reconstructed to decode the current EL, and decoding complexity is higher than that of the single-loop decoding design in SVC [14]. More detailed reviews of the SHVC standard and HEVC extensions can be found in [15]–[17].

3 SHVC in ATSC 3.0

3.1 ATSC 3.0

In the past, delivery of video entertainment to the consumer was relatively simple and controlled, and involved broadcasters or content producers sending TV signals at prescheduled times to the living room. Today, people watch video on-demand on a wide variety of devices at a time and place of their choice. The delivery paths may be over-the-air, cable or satellite, Internet, local storage, or a combination of these. ATSC 3.0 is the next-generation broadcast standard designed to address this need. It uses advanced transmission, including hybrid broadcast and broadband, as well as advanced video/audio coding techniques to bring new, creative services to viewers [18].

The next-generation ATSC 3.0 broadcast system is designed to increase service flexibility and enable terrestrial broadcasters to send hybrid-content services to fixed and mobile receivers in a seamless manner. It combines both over-the-air transmission and broadband delivery. Other essential features include support for multiscreen and the flexibility to choose among SD, HD and UHD resolutions. SHVC provides an efficient solution when different spatial resolutions need to be transmitted by the content provider at the same time.

The work on ATSC 3.0 is organized according to layers, such as the physical layer, management and protocol layer, and application and presentation layer. Video coding, audio coding, and run-time environment are addressed by the application and presentation layer. Support for UHD and HD is key for video coding — 4K support at the start and potentially 8K support via future extensions. Portable, handheld, vehicular, and fixed devices (both indoors and outdoors) are all targeted, and hybrid integration of broadcast and broadband delivery is required. This paper mainly focuses on the work by ATSC S34-1, the ad hoc group for video for ATSC 3.0. A general overview of ATSC 3.0 and all ATSC 3.0 groups can be found in [19].

3.2 SHVC Performance with ATSC 3.0

Video requires support for UHD and HD; support for portable, mobile, vehicular, and fixed devices operating in indoors or outdoors; and support for hybrid broadcast/broadband delivery. The following four scenarios were identified for ATSC 3.0 deployment:

1) larger coverage area (scenario A). Receivers in a first class are fixed within the current ATSC 1.0 coverage area, and receivers in a second class are fixed but are not within the coverage area (e.g., rural, or with an indoor or integrated antenna).
2) pedestrian phone or tablet (scenario B). Receivers in a first class are handheld and moving at pedestrian speeds (possibly indoors), and receivers in a second class are stationary.
3) mobile-enabled (scenario C). Receivers in a first class are moving at relatively high speed, and receivers in a second class are stationary.
4) tablet in bedroom (scenario D). Receivers in a first class are indoors and are portable, and receivers in a second class are stationary.

SHVC was evaluated in each of these four cases, with the main focus on spatial scalability; i.e., the base layer could be optimized for mobile reception and the enhancement layer could be optimized for up to 4K resolution. These four scenarios were proposed and agreed upon by S34-1 to be used as common test conditions for comparing the performance of SHVC with HEVC simulcast. In each scenario, three different physical-layer pipes (PLPs)—PLP-1, PLP-2 and PLP-3—were assumed for transmitting high-quality video, low-quality video, and audio (and miscellaneous information), respectively. The video resolution, spectral efficiency, and coded bit rate for each PLP and each scenario are summarized in **Table 1**. In all four scenarios, the sum of bandwidths of all PLPs (after spectral efficiency has been taken into account) does not exceed 6 Mbps. The detailed test conditions can be found in [19].

To make a meaningful comparison, both HEVC and SHVC

Table 1. ATSC 3.0 common test conditions for SHVC

Scenario	Configuration	PLP-1	PLP-2	PLP-3
A	Resolution	UHD (2160)	HD (1080)	Audio/misc.
	Spectral Efficiency (b/s/Hz)	4.0	2.67	1.31
	Bit rate (Mbps)	15.1	5.0	0.47
B	Resolution	HD (1080)	HD (720)	Audio/misc.
	Spectral Efficiency (b/s/Hz)	2.23	1.0	1.0
	Bit rate (Mbps)	5.0	3.46	0.3
C	Resolution	HD (1080)	qHD (540)	Audio/misc.
	Spectral Efficiency (b/s/Hz)	4.0	0.44	0.44
	Bit rate (Mbps)	5.0	1.7	0.34
D	Resolution	UHD	HD	Audio
	Spectral Efficiency (b/s/Hz)	7.1	0.59	0.44
	Bit rate (Mbps)	4.5	2.75	0.3

HD: high definition　qHD: quarter high definition
PLP: physical-layer pipe　UHD: ultra-high definition

were coded using configurations that were as similar as possible. The hierarchical B configuration following the SHVC common test condition in JCT-VC [10] was used, with the random access point period of 0.5 seconds for the BL and 0.5 seconds or 4 seconds for the EL. The quality range was controlled by using a PSNR of between 38 dB and 42 dB, the typical operating quality for broadcasters. The bit rates of both layers were controlled so that they were no higher than those listed in Table 1.

Fig. 2 shows the performance of SHVC and HEVC simulcast for each scenario. Two types of savings were calculated: 1) the percentage of average video bit rate savings, which is how much SHVC reduces the coded video bit rate compared to simulcast while maintaining the same quality (PSNR); and 2) the percentage of average channel utilization savings, which is calculated by converting bit rate savings into actual channel utilization savings by taking into account the different spectral efficiencies for each PLP.

In general, SHVC provides 40%–47% video bit rate savings in the four scenarios and 6%–37% channel utilization savings in the four scenarios when spectral efficiency is taken into account. Channel utilization is inversely proportional to the spectral efficiency factors in Table 1. In the ATSC tests, the BL bit rates were fixed; therefore, the spectral efficiency for the BL (PLP-2) does not have any effect. A bigger spectral efficiency factor for the EL (PLP-1) will translate the same amount of bit rate saving into less channel utilization saving. This is why the channel utilization savings for scenario D with the PLP-1 spectral efficiency of 7.1 is significantly less than that of scenario A with PLP-1 spectral efficiency of 4.0. The detailed performance comparison can be found in [20].

4 SHVC in 3GPP SA4

4.1 3GPP SA4

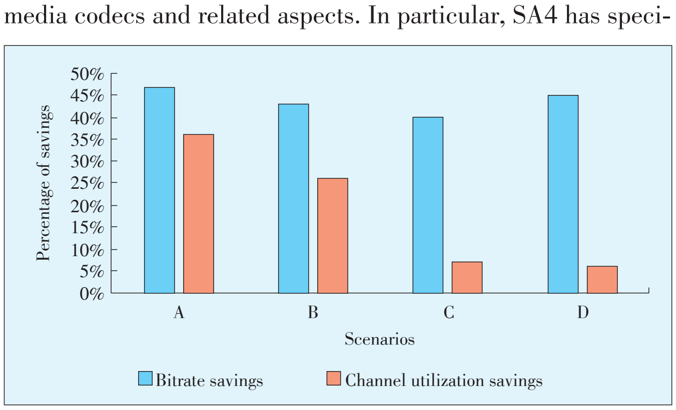
SA4 is the 4th working group of the 3GPP Technical Specification Group of Service and System Aspects (TSG-SA). SA4 is responsible for development of 3GPP standards that handles media codecs and related aspects. In particular, SA4 has specified the media handling aspects of all 3GPP multimedia service standards, including 3GP-DASH in TS 26.247 [21], Packet-switched Streaming Service (PSS) in TS 26.234 [22], MBMS in TS 26.346 [23], Multimedia Telephony Service over IMS (MTSI) in TS 26.114 [24] (this also includes MMVC), Multimedia Messaging Service (MMS) in TS 26.140 [25], IMS Messaging and Presence in TS 26.141 [26], and IMS based Telepresence in TS 26.223 [27].

For each of these multimedia services, the selection of media codecs to be supported is important. Support for H.264/AVC in 3GPP multimedia services was decided in 2004, e.g., it was first included in TS 26.234 in v6.1.0 dated in September 2004. SVC was studied in 2010 and the result was included in TR 26.904 [28]; it was decided not to specify SVC support in the 3GPP multimedia services. For HEVC, a specific work item was agreed by SA4 in August 2012, and a study was performed comparing HEVC with H.264/AVC and documented in TR 26.906 [29]. For SHVC, a study item was agreed by SA4 in November 2014, focusing on evaluation of SHVC versus HEVC simulcast for three of the 3GPP multimedia services: the MMVC part of MTSI, MBMS, and 3GP-DASH. The use cases and simulations for MMVC happened to apply to the telepresence service. This study was completed in November 2015. An overview of the SHVC use cases, simulation results, and complexity analyses is provided in the following subsections.

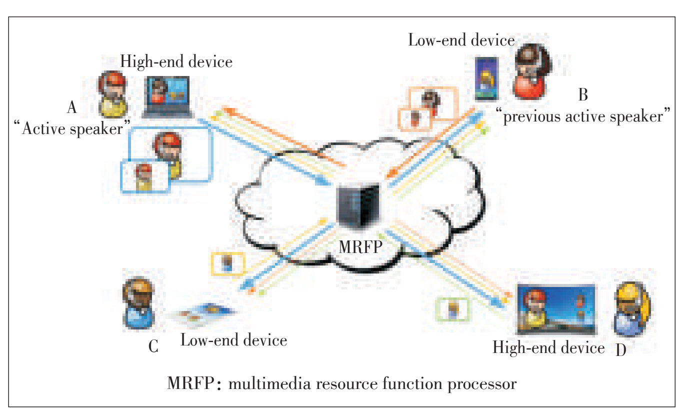
4.2 Using SHVC for MMVC and Telepresence

The performance of SHVC was evaluated for the MMVC and telepresence use cases in 3GPP SA4 [30]. The use case considers video conferencing with multiple participating user equipment (UE) with different decoding and display capabilities. The multimedia resource function processor (MRFP) connects multiple video conferencing endpoints, receives video streams from each endpoint, and forwards a set of appropriate video streams to each endpoint.

Fig. 3 illustrates an example of such use case with four UEs in the video conferencing session, where UE-A and UE-D are high-end devices and UE-B and UE-C are low-end devices. Each UE displays a full video of the active speaker and a num-



▲Figure 2. SHVC vs. HEVC simulcast.



▲Figure 3. The use case for MMVC and telepresence.

ber of thumbnails of all other participants, while the active speaker displays the previous active speaker as full video. In Fig. 3, UE-A is the current active speaker sending a high video resolution and a medium video resolution to the MRFP. The high-resolution video is forwarded to participants with a high-end device (UE-D), and the medium resolution video is forwarded to participants with a low-end device (UE-B and UE-C). The current active speaker (UE-A) receives the medium resolution video from the previous active speaker (UE-B). Each UE except the active speaker receives either high-resolution or medium-resolution video of the active speaker from the MRFP and sends a low-resolution thumbnail video to the MRFP to be displayed by other UEs.

For HEVC simulcast, on the uplink side, UE-A sends one high-resolution video bitstream and one medium-resolution video bitstream to the MRFP, UE-B sends a medium-resolution video bitstream and a thumbnail video bitstream to the MRFP, and UE-C and UE-D each sends a thumbnail video to the MRFP. On the downlink side, each UE except the active speaker receives one high- or medium-resolution video bitstream of the active speaker depending on the device capability for full video display, and a thumbnail video bitstream from each of the other UEs for thumbnail display. The active speaker receives the medium-resolution video bitstream of the previous active speaker for full video display, and a thumbnail video bitstream from each of the other UEs for thumbnail display.

For SHVC, on the uplink side, UE-A sends a two-layer SHVC bitstream with BL at medium resolution and EL at high resolution to the MRFP. UE-B sends a two-layer SHVC bitstream with BL at thumbnail resolution and EL at medium resolution to the MRFP. UE-C and UE-D each sends an HEVC single layer thumbnail video bitstream to the MRFP. On the downlink side, UE-A receives a two-layer SHVC bitstream from UE-B for full video display, and two HEVC single-layer bitstreams from UE-C and UE-D for thumbnail display. UE-B receives the extracted BL bitstream from UE-A for full video display, and two HEVC bitstreams from UE-C and UE-D for thumbnail display. UE-C receives one extracted BL bitstream from UE-A for full video display, one extracted BL bitstream from UE-B for thumbnail display, and one HEVC single-layer bitstream from UE-D for thumbnail display. UE-D receives one two-layer SHVC bitstream from UE-A for full video display, one extracted BL bitstream from UE-B for thumbnail display, and one HEVC single-layer bitstream from UE-C for thumbnail display.

In the simulations, the high video resolution was 1080p, the medium video resolution was 720p, and thumbnail video resolution was 240p.

Table 2 shows the SHVC rate savings on the uplink and rate penalty on the downlink, for each participant. On the uplink, UE-A saves on average 27.3% bandwidth, and UE-B saves on average 5.5% bandwidth. On the downlink, because the two-layer bitstream needs to be received when SHVC is used, UE-A's downlink bandwidth increases by 11.6%, UE-D's downlink bandwidth increases by 23.5%. For UE-C and UE-D, the uplink bandwidth usage is identical, regardless of the codec choice. The same is true for downlink bandwidth usage for UE-B and UE-C.

▼**Table 2. Uplink and downlink rate saving comparison for MMVC/telepresence**

	UE-A	UE-B	UE-C	UE-D
Average uplink bandwidth saving	27.3%	5.5%	0%	0%
Average downlink bandwidth cost	11.6%	0%	0%	23.5%

UE: user equipment

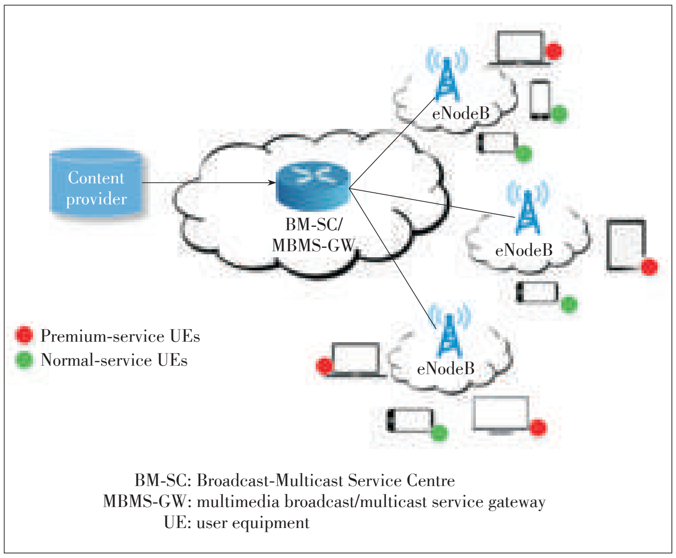
In general, SHVC provides uplink bandwidth savings for UEs that are sending more than one video resolution, and incurs downlink bandwidth penalty for UEs that are receiving the high-resolution video. Further detailed results can be found in [30]–[33].

4.3 Using SHVC for MBMS

The MBMS case is referred to as the differentiated-service MBMS use case [33]. For this use case, it is assumed that two different classes of video services may be provided (more classes of video service is possible but could pose burden for any broadcast system), e.g., the normal video service of 720p and the premium video service of 1080p. UEs may subscribe to either of the two services depending on their decoding and rendering capabilities, network access conditions, power saving strategies, price, and/or other considerations. UEs receiving the normal service receive and render the lower quality video with lower resolution, and UEs receiving the premium service receive and render the higher quality video with higher resolution. The same scenario is also applicable to evolved MBMS (eMBMS), which allows broadcast over the LTE network. Due to the fact that only the broadcast mode can be used in eMBMS, all bits required for both services are assumed to be transmitted on all the network paths, from the content provider to the Broadcast-Multicast Service Centre (BM-SC), from the BM-SC to MBMS Gateway (MBMS-GW), from MBMS-GW to evolved Node B (eNodeB), as well as the air interface between eNodeB and UEs, as shown in **Fig. 4**.

When SHVC is used in the differentiated-service MBMS use-case, the content is encoded with two layers of different spatial resolutions, and is transmitted from the content provider to the BM-SC, and all the way to the UEs. Each premium-service UE receives and decodes both layers and renders the higher layer, while each normal-service UE receives, decodes, and renders the base layer only.

The performance of SHVC was evaluated for the MBMS use case against HEVC simulcast. Five test sequences with 720p for the BL and 1080p for the EL were used. For HEVC simulcast, the bandwidth for transmission from the content provider to the BM-SC, and all the way to the UEs is the bandwidth required for transmitting one HEVC coded 1080p bitstream and



▲Figure 4. The use case for MBMS.

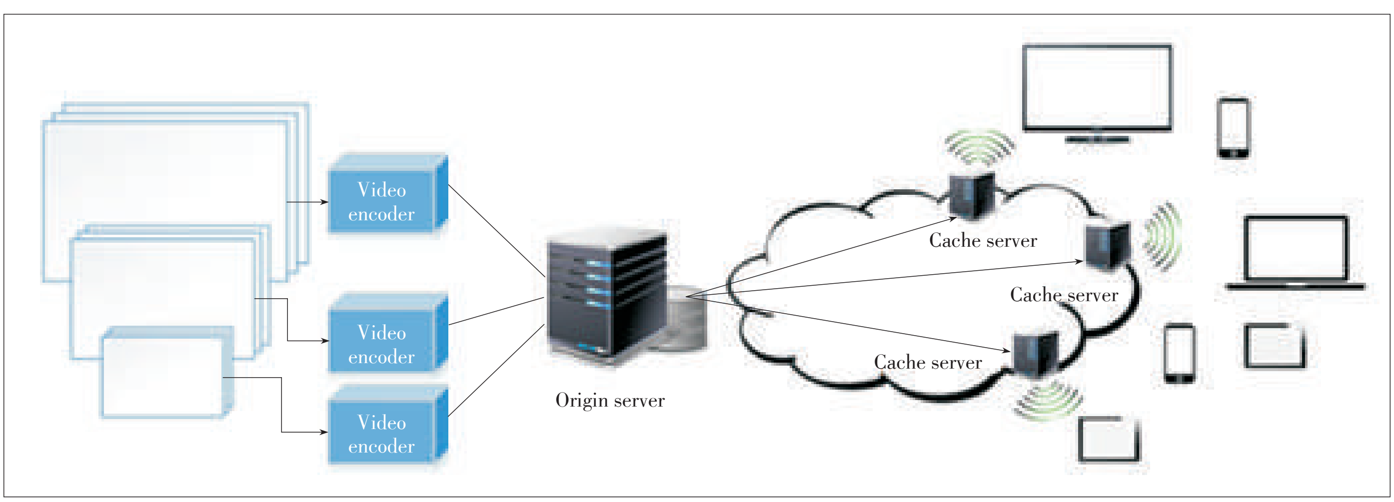
one HEVC coded 720p bitstream. In the simulations, video bitstreams were encoded with a random access coding structure to achieve the highest compression efficiency. Furthermore, to enable stream switching or late tuning-in and channel switching in MBMS, intra random access point (IRAP) picture is coded once every two seconds. Further details of the simulation condition can be found in [33]. The performance of SHVC compared to HEVC simulcast for the MBMS use case in terms of bandwidth reduction, decoding complexity and encoding complexity are summarized as follows (further details can be found in [33], [34]):

1) In term of bandwidth reduction, the use of SHVC provides an average bandwidth reduction around 32.9% when compared to HEVC simulcast.
2) The decoding complexity overhead at UEs depends on how many layers an UE needs to decode. The decoding complexity for UEs receiving normal-service when SHVC is used can be assumed the same as when HEVC simulcast is used because UEs receiving normal-service can ignore coded data for enhancement layer. The decoding complexity overhead for UEs receiving the premium-service when SHVC is used is roughly the percentage of the number of samples in the lower resolution video relative to that in the higher resolution video.
3) Compared with simulcast, SHVC encoding may be less complex than simulcast encoding because SHVC places the zero-motion constraint on inter layer prediction. When the inter-layer reference picture provides a sufficiently good prediction signal (without the need for motion estimation), early termination is typically applied at the encoder, and the need for motion estimation of the temporal reference pictures is avoided, leading to lower encoding complexity.

4.4 Using SHVC for the 3GP-DASH Use Case

The use case scenario the 3GP-DASH video streaming services involves a diverse of end user devices which could have different display capabilities and network access conditions [33]. Each UE may prefer to receive a different quality of content, possibly with a different resolution, and request the chosen video content from the origin server, involving cache servers between the origin server and the UE. During a session, an UE may also adaptively switch to segments of different representations of different bit rates and qualities and possibly also different spatial resolutions to adapt to the dynamic network conditions. Video content is encoded into multiple video streams in different representations providing different levels of resolutions or qualities, e.g., as three representations of resolutions 360p, 720p and 1080p (**Fig. 5**). Copies of the streams may be stored in the cache servers and directly served to the UEs.

When SHVC is used, multiple resolutions or quality representations can be encoded into multi-layer SHVC bitstreams. Each layer can be encapsulated as one 3GP-DASH representa-



▲Figure 5. The use case for 3GP-DASH.

tion. A client wanting a particular resolution or quality can request segments of that representation and all other representations it depends on (i.e., request the desired layer and all layers the desired layer depends on). The desired layer and all its dependent layers are then sent to the client, which decodes the bitstream and outputs the desired layer.

The performance of SHVC was evaluated for the 3GP-DASH use case against HEVC simulcast. The simulations were conducted with three representations of spatial resolution 360p, 720p and 1080p, the random access coding structure, and one IRAP picture every two and four seconds. Further details of the simulation condition can be found in [33]. The performance of SHVC compared to HEVC simulcast from the aspects of required bandwidth for transmission, decoding complexity and encoding complexity are as follows (further details can be found in [33] and [35]):

1) For outgoing transmission bandwidth, i.e., bandwidth required for transmission of encoded content from the origin server to cache servers and from the origin server to UEs, compared to HEVC simulcast SHVC requires less bandwidth for transmitting the encoded streams from the origin server to cache and to UEs. The bandwidth reduction varies from 9.2% to 10.5% for transmitting both the 360p and 720p bitstreams and from 23.3% to 23.6% for transmitting all the 360p, 720p and 1080p bitstreams. In addition to saving the outgoing bandwidth, the same amount of savings can be achieved on the storage requirements for the origin server and the cache servers. For incoming transmission bandwidth, i.e., bandwidth required by UEs to receive the encoded content, SHVC incurs data overhead for UEs when receiving the medium or high resolution representation. The overhead varies from 20.4% to 22.1% when receiving the 720p resolution and from 24.9% to 26.9% when receiving the 1080p.
2) The decoding complexity is mainly proportional to the resolution(s) of the video represented in the bitstream. For HEVC simulcast, only one single layer stream needs to be decoded, i.e., one of the three bitstreams of 360p, 720p and 1080p. For SHVC, the decoding complexity depends on the resolution of each layer that needs to be decoded in order to output the highest layer video resolution.
3) For HEVC simulcast, the content provider has to encode independent bitstreams of different spatial resolutions. For SHVC, the content provider has to encode a bitstream with multiple layers in which each layer is associated with one spatial resolution. Compared to simulcast, the complexity of SHVC encoding may be less than that of simulcast encoding for the same reason as discussed in the MBMS use case.

5 Conclulsions

In this paper, a brief overview of SHVC, the latest scalable video coding standard based on HEVC, was provided. Several use cases for SHVC, as were recently studied by application SDOs including ATSC and 3GPP SA4, were reviewed. In the broadcasting and multicasting cases, SHVC saves transmission bandwidth. In the video conferencing and telepresence cases, SHVC saves uplink bandwidth but increases the downlink rate for high-end devices. In the DASH-based video streaming case, SHVC saves server storage and outgoing transmission bandwidth but increases incoming transmission bandwidth for devices receiving representations with higher bit rates, picture rates, spatial resolutions and so on. The decoding complexity for clients processing an SHVC bitstream is higher than that for clients processing a corresponding HEVC bitstream in simulcast, whereas the encoding complexity is typically lower. SHVC was recently included in the ATSC 3.0 standard based on the significant channel utilization savings it can provide for the broadcasters. 3GPP concluded that SHVC can provide technical benefits in different scenarios and circumstances and may be an attractive codec solution whenever new use cases and scenarios are considered within emerging 3GPP multimedia services. However, a normative specification of SHVC support in a 3GPP Release 13 multimedia service standard has not been included.

References

[1] *High Efficiency Video Coding*, Rec. ITU-T H.265 and ISO/IEC 23008-2, Version 1-3, Apr. 2013-Apr. 2015.
[2] *Advanced Video Coding for Generic Audiovisual Services*, Rec. ITU-T H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), Version 8, Jul. 2007.
[3] T. K. Tan, M. Mrak, V. Baroncini, and N. Ramzan, "Report on HEVC compression performance verification testing," Joint Collaborative Team on Video Coding, JCTVC-Q1011, Valencia, Spain, 27 March-4 April 2014.
[4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012. doi: 10.1109/TCSVT.2012.2221191.
[5] G. J. Sullivan and J.-R. Ohm, "Joint call for proposals on scalable video coding extensions of high efficiency video coding (HEVC)," ITU-T Study Group 16 Video Coding Experts Group (VCEG) document VCEG-AS90 and ISO/IEC JTC 1/SC 29/WG 11 (MPEG) document N12957, Jul. 2012.
[6] J. Chen, J. Boyce, Y. Ye, *et al.*, "High efficiency video coding (HEVC) scalable extension draft 7," Joint Collaborative Team on Video Coding, JCTVC-R1008_v7, Sapporo, Japan, 2014.
[7] J. Boyce, J. Chen, Y. Chen, *et al.*, "Draft high efficiency video coding (HEVC) version 2, combined format range extensions (RExt), scalability (SHVC), and multi-view (MV-HEVC) extensions," Joint Collaborative Team on Video Coding, JCTVC-R1013_v6, Sapporo, Japan, 2014.
[8] J. Chen, J. Boyce, Y. Ye, and M. M. Hannuksela, "SHVC Test Model 10 (SHM 10) Introduction and Encoder Description," Joint Collaborative Team on Video Coding, JCTVC-U1007, Warsaw, Poland, Jun. 2015.
[9] SHM-10.0 Reference Software [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/tags/SHM-10.0
[10] V. Seregin and Y. He, "Common SHM test conditions and software reference configurations," JCT-VC, Joint Collaborative Team on Video Coding (JCT-VC) Document JCTVC-Q1009, Valencia, Spain, Ap. 2014.
[11] *Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-*

Screen 16:9 Aspect Ratios, ITU-R Recommendation BT.601, 1982.
[12] *Parameter Values for the HDTV Standards for Production and International Programme Exchange*, ITU-R Recommendation BT.709, Dec. 2010.
[13] *Parameter Values for UHDTV Systems for Production and International Programme Exchange*, ITU-R Recommendation BT.2020, Aug. 2012.
[14] H. Schwarz, D. Marpe, and T. Wiegand, "Constrained inter-layer prediction for single-loop decoding in spatial scalability," in *Proc. ICIP*, Genoa, Italy, Sep. 2005, vol. 2, pp. 870-873. doi: 10.1109/ICIP.2005.1530194.
[15] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: the scalable extensions of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20-34, Jan. 2016. doi: 10.1109/TCSVT.2015.2461951.
[16] Y. Ye and P. Andrivon, "The scalable extensions of HEVC for ultra high definition video delivery," *IEEE Transactions on Circuits and Systems for Video Technology*, to appear.
[17] G. J. Sullivan, J. M. Boyce, Y. Chen, *et al.*, "Standardized extensions of high efficiency video coding (HEVC)," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001 - 1016, Dec. 2013. doi: 10.1109/JSTSP.2013.2283657.
[18] ATSC. ATSC 3.0: Where We Stand [Online]. Available: http://atsc.org/newsletter/atsc-3-0-where-we-stand
[19] ATSC, "S34-1-123r3-On Common Conditions for SHVC," ATSC 34-1 document, Jan. 2015.
[20] ATSC, "S34-1-169r1-Summary of SHVC Testing," ATSC 34-1 document, May 2015.
[21] *Transparent End-to-End Packet-Switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)*, 3GPP TS 26.247, 2015.
[22] *Transparent End-to-End Packet-Switched Streaming Service (PSS); Protocols and Codecs*, 3GPP TS 26.234, 2015.
[23] *Multimedia Broadcast/Multicast Service (MBMS); Protocols and Codecs*, 3GPP TS 26.346, 2015.
[24] *IP Multimedia Subsystem (IMS); Multimedia Telephony; Media Handling and Interaction*, 3GPP TS 26.114, 2015.
[25] *Multimedia Messaging Service (MMS); Media Formats and Codecs*, 3GPP TS 26.140, 2015.
[26] *IP Multimedia System (IMS) Messaging and Presence; Media Formats and Codecs*, 3GPP TS 26.141, 2015.
[27] *Telepresence Using the IP Multimedia Subsystem (IMS); Media Handling and Interaction*, 3GPP TS 26.223, 2015.
[28] *Improved Video Coding Support*, 3GPP TR 26.904, 2015.
[29] *Evaluation of High Efficiency Video Coding (HEVC) for 3GPP Services*, 3GPP TR 26.906, 2015.
[30] *FS_VE_3MS: Use Case for MMVC*, 3GPP SA4, S4-150966, Aug. 2015.
[31] *FS_VE_3MS: Additional Simulation Results for MMVC and Telepresence*, 3GPP SA4, S4-151317, Oct. 2015.
[32] *FS_VE_3MS: Additional Cost and Benefit Comparison Between SHVC and HEVC Simulcast for MMVC and Telepresence Use Cases*, 3GPP SA4, S4-151318, Oct. 2015.
[33] *Study on Video Enhancements in 3GPP Multimedia Services*, 3GPP TR 26.948 V1.2.0, 2015.
[34] *FS_VE_3MS: Use Case for MBMS*, 3GPP SA4, S4-150967, Aug. 2015.
[35] *FS_VE_3MS: Use-Case for 3GP-DASH*, 3GPP SA4, S4-150968, Aug. 2015.

Manuscript received: 2015-11-19

Biographies

Yan Ye (Yan.Ye@InterDigital.com) received her PhD from the Electrical and Computer Engineering Department at University of California, San Diego in 2002. She received her MS and BS degrees, both in Electrical Engineering, from the University of Science and Technology of China, in 1994 and 1997, respectively. She is currently with the Wireless Business Unit Labs at InterDigital Communications as a Senior Manager, where she manages the video standards and platforms project. Previously she was with Image Technology Research at Dolby Laboratories Inc and Multimedia R&D and Standards at Qualcomm Inc. She has been involved in the development of various international video coding standards, including the HEVC standard, the scalable extensions, screen content coding extensions, and high dynamic range extensions of HEVC, and the scalable extensions of H.264/AVC. Her research interests include video coding, processing and streaming. She is the co-inventor of 40+ granted US patents. She received InterDigital's Innovation Awards and Publication Award in 2012, 2013 and 2014.

Yong He (Yong.He@InterDigital.com) received his PhD degree from Hong Kong University of Science and Technology, MS and BS degrees from Southeast University, Nanjing, China. He is currently a member of technical staff in InterDigital Communications, Inc, San Diego, CA, USA. His early working experiences and titles include Principal Staff Engineer at Motorola, San Diego, CA, USA, from 2001 to 2011, and Research Engineer at Motorola Australia Research Center, from 1999 to 2001. He is currently active in video related standardization and platform development. He has co-authored various technical standardization contributions, academic papers, and over 10 granted US patents. His research interests include video coding, analysis and processing.

Ye-Kui Wang (yekuiw@qti.qualcomm.com) received his BS degree in industrial automation in 1995 from Beijing Institute of Technology, and his PhD degree in electrical engineering in 2001 from the Graduate School in Beijing, University of Science and Technology of China. He is currently a Director of Technical Standards at Qualcomm, San Diego, CA, USA. Previously, he worked at Huawei Technologies, Nokia Corporation, and Tampere University of Technology. His research interests include video coding and multimedia transport and systems. He has been a contributor to various multimedia standards of video codecs, file formats, RTP payload formats, HTTP streaming, and video application systems, and an editor of several standards, including HEVC, SHVC, HEVC file format, HEVC RTP payload format, SHVC/MV-HEVC file format, H.271, SVC file format, MVC, RFC 6184, RFC 6190, and 3GPP TR 26.906 and 26.948. He has co-authored over 500 standardization contributions, about 50 academic papers, and over 200 families of granted or pending patents.

Hendry (hendry@qti.qualcomm.com) received his PhD and MS degrees from Korea Advanced Institute of Science and Technology, in 2011 and 2005, respectively, BS degree from University of Indonesia, in 2002. He is currently a staff engineer in Qualcomm, Inc, San Diego, CA, USA. His early working experience and title includes Senior Engineer at LG Electronics, Seoul, Korea, from 2011 to 2013. He has been involved in the development of MPEG standards including MPEG-21, MPEG Multimedia Applications Formats, MPEG-2 Transport Systems and ISO-Based File Format and its derivations, video coding standards, including the HEVC standard and its multi-layer extensions. He is the co-inventor of 30+ filed US patents. His research interests include video coding and its transport systems.