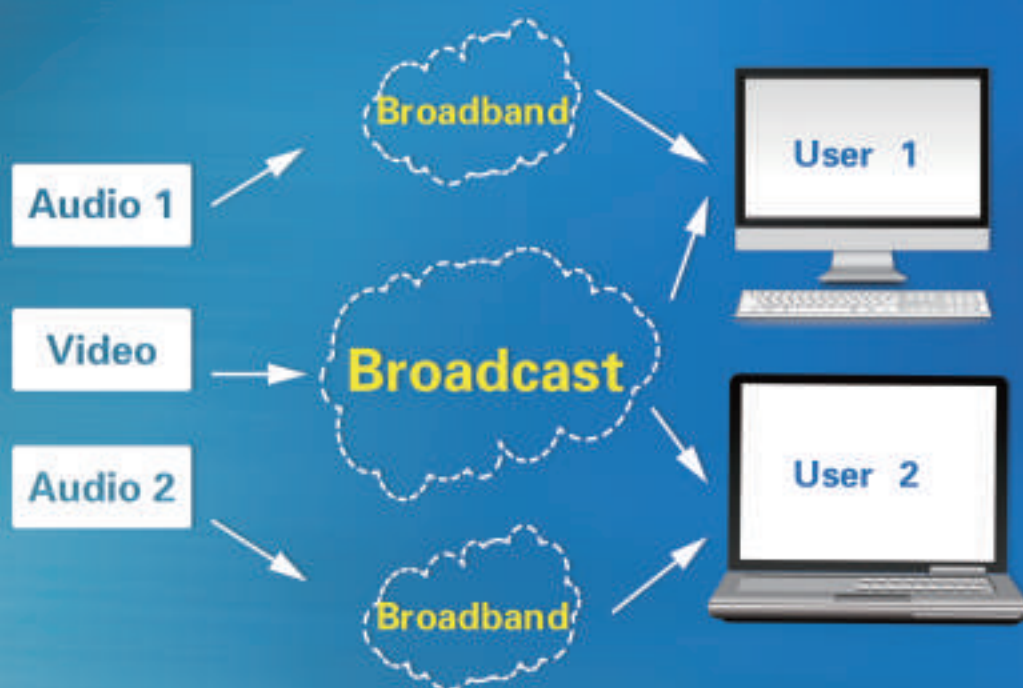


# ZTE COMMUNICATIONS

An International ICT R&D Journal Sponsored by ZTE Corporation

February 2016, Vol. 14 No. 1

## SPECIAL TOPIC: Emerging Technologies of Future Multimedia Coding, Analysis and Transmission



# ZTE Communications Editorial Board

## Chairman

**Houlin Zhao:** International Telecommunication Union (Switzerland)

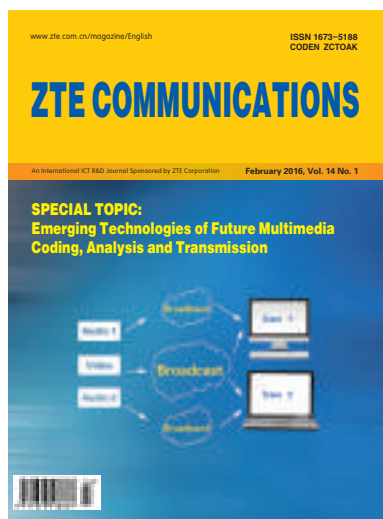
## Vice Chairmen

**Lirong Shi:** ZTE Corporation (China)    **Chengzhong Xu:** Wayne State University (USA)

## Members (in Alphabetical Order):

Chang Wen Chen	The State University of New York at Buffalo (USA)
Chengzhong Xu	Wayne State University (USA)
Connie Chang-Hasnain	University of California, Berkeley (USA)
Fa-Long Luo	Element CXI (USA)
Fuji Ren	The University of Tokushima (Japan)
Guifang Li	University of Central Florida (USA)
Honggang Zhang	Université Européenne de Bretagne (France)
Houlin Zhao	International Telecommunication Union (Switzerland)
Huifang Sun	Mitsubishi Electric Research Laboratories (USA)
Jianhua Ma	Hosei University (Japan)
Jiannong Cao	Hong Kong Polytechnic University (Hong Kong, China)
Jie Chen	ZTE Corporation (China)
Jinhong Yuan	University of New South Wales (Australia)
Keli Wu	The Chinese University of Hong Kong (Hong Kong, China)
Kun Yang	University of Essex (UK)
Lirong Shi	ZTE Corporation (China)
Shigang Chen	University of Florida (USA)
Shuguang Cui	Texas A&M University (USA)
Victor C. M. Leung	The University of British Columbia (Canada)
Wanlei Zhou	Deakin University (Australia)
Weihua Zhuang	University of Waterloo (Canada)
Wen Gao	Peking University (China)
Wenjun (Kevin) Zeng	University of Missouri (USA)
Xiaodong Wang	Columbia University (USA)
Yi Pan	Georgia State University (USA)
Yingfei Dong	University of Hawaii (USA)
Yueping Zhang	Nanyang Technological University (Singapore)
Zhili Sun	University of Surrey (UK)

# ► CONTENTS



Submission of a manuscript implies that the submitted work has not been published before (except as part of a thesis or lecture note or report or in the form of an abstract); that it is not under consideration for publication elsewhere; that its publication has been approved by all co-authors as well as by the authorities at the institute where the work has been carried out; that, if and when the manuscript is accepted for publication, the authors hand over the transferable copyrights of the accepted manuscript to *ZTE Communications*; and that the manuscript or parts thereof will not be published elsewhere in any language without the consent of the copyright holder. Copyrights include, without spatial or timely limitation, the mechanical, electronic and visual reproduction and distribution; electronic storage and retrieval; and all other forms of electronic publication or any other types of publication including all subsidiary rights.

Responsibility for content rests on authors of signed articles and not on the editorial board of *ZTE Communications* or its sponsors.

All rights reserved.

## Special Topic: Emerging Technologies of Future Multimedia Coding, Analysis and Transmission

### Guest Editorial

01

Huifang Sun, Can Shen, and Ping Wu

### Overview of the Second Generation AVS Video Coding Standard (AVS2)

03

Shanshe Wang, Falei Luo, and Siwei Ma

### An Introduction to High Efficiency Video Coding Range Extensions

12

Bin Li and Jizheng Xu

### Multi-Layer Extension of the High Efficiency Video Coding (HEVC) Standard

19

Ming Li and Ping Wu

### SHVC, the Scalable Extensions of HEVC, and Its Applications

24

Yan Ye, Yong He, Ye-Kui Wang, and Hendry

### ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution

32

Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fröhlich, and Guan-Ming Su

# ▶ CONTENTS

## ZTE COMMUNICATIONS

Vol. 14 No. 1 (Issue 49)

Quarterly

First English Issue Published in 2003

### Supervised by:

Anhui Science and Technology Department

### Sponsored by:

Anhui Science and Technology Information  
Research Institute and ZTE Corporation

### Staff Members:

Editor-in-Chief: Jie Chen

Executive Associate

Editor-in-Chief: Huang Xinming

Editor-in-Charge: Zhu Li

Editors: Paul Sleswick, Xu Ye, Lu Dan,  
Zhao Lu

Producer: Yu Gang

Circulation Executive: Wang Pingping

Assistant: Wang Kun

### Editorial Correspondence:

Add: 12F Kaixuan Building,

329 Jinzhai Road,

Hefei 230061, P. R. China

Tel: +86-551-65533356

Fax: +86-551-65850139

Email: magazine@zte.com.cn

### Published and Circulated

(Home and Abroad) by:

Editorial Office of

*ZTE Communications*

### Printed by:

Hefei Tiancai Color Printing Company

### Publication Date:

February 25, 2016

### Publication Licenses:

ISSN 1673-5188

CN 34-1294/TN

### Advertising License:

皖合工商广字0058号

### Annual Subscription:

RMB 80

## DASH and MMT and Their Applications in ATSC 3.0

39

Yiling Xu, Shaowei Xie, Hao Chen, Le Yang, and Jun Sun

## Introduction to AVS2 Scene Video Coding Techniques

50

Jiaying Yan, Siwei Dong, Yonghong Tian, and Tiejun Huang

## Review

## From CIA to PDR: A Top-Down Survey of SDN Security for Cloud DCN

54

Zhi Liu, Xiang Wang, and Jun Li

## Research Papers

## A Software-Defined Approach to IoT Networking

61

Christian Jacquenet and Mohamed Boucadair

## Roundup

## Call for Papers: Special Issue on Multiple Access Techniques for 5G

02

## Call for Papers: Special Issue on Multi-Gigabit Millimeter-Wave Wireless Communications

18

## Introduction to *ZTE Communications*

66

# Emerging Technologies of Future Multimedia Coding, Analysis and Transmission

## ► Huifang Sun



Huifang Sun (hsun@merl.com) received his PhD degree from the University of Ottawa, Canada. In 1990, he was an associate professor at Fairleigh Dickinson University. Also in 1990, he joined Sarnoff Corporation as a member of the technical staff and was later promoted to technology leader. In 1995, he joined Mitsubishi Electric Research Laboratories and was promoted to vice president, deputy director, and fellow (2003). He has co-authored two books and published more than 140 journal and conference papers. He holds more than 60 US patents. In 1994, Dr. Sun received a Technical Achievement Award for optimization and specification of the Grand Alliance HDTV video compression algorithm. In 1992, he won the Best Paper award from IEEE Transaction on Consumer Electronics. In 1996, he won the Best Paper award at ICCE, and in 2003, he won the Best Paper award from IEEE Transactions on CSVT. He has been associate editor of IEEE Transaction on Circuits and Systems for Video Technology and was the chair of the Visual Processing Technical Committee of IEEE's Circuits and System Society. He is an IEEE Fellow.

## ► Can Shen



Can Shen (shen.can1@zte.com.cn) received his PhD degree in physical electronics from Southeast University, China in 1997. From 1997 to 2000, he was a lecturer in Nanjing University of Posts and Telecommunications. From 2000, he joined ZTE as a senior engineer and was later promoted to the chief engineer. He has published more than 15 papers and holds more than 50 patents.

## ► Ping Wu



Ping Wu (ping.wu@zte.com.cn) received his PhD degree in signal processing from Reading University, United Kingdom in 1993. From 1993 to 1997, he was a Research Fellow in the area of medical data processing in Plymouth University, United Kingdom. From 1997 to 2008, he was a Consultant Engineer in News Digital Systems Ltd, Tandberg Television, and Ericsson. He participated in the development of ISO/IEC MPEG and ITU-T video coding standards. He also supervised the engineering team to build the High Definition H.264 encoder products for broadcasters. From 2008 to 2011, he joined Mitsubishi Electric Research Centre Europe and continued to participate in HEVC standard development with contributions in Call for Evidence and Call for Proposal. From 2011, he has been a senior specialist in video coding in ZTE. He has many technical proposals and contributions to the international standards on video coding over past 18 years.

Three years ago, *ZTE Communications* published a special issue called Emerging Technologies of Multimedia Coding, Analysis and Transmission. Over the past three years, great advances have been made in multimedia. The purpose of this special issue is to report on the progress and achievements in this field. We invited 11 papers, seven of which will be published in this issue and four in the next issues as research papers. All authors we invited are the top researchers in the area of multimedia from both academic and industry. Also, these papers have been reviewed by the experts who are working in the front of this area.

The paper "Overview of the Second Generation AVS Video Coding Standard (AVS2)" by Shanshe Wang *et al.* introduces a new generation video coding standard developed by the AVS working group. Compared with the first generation video coding standard AVS1, AVS2 significantly improves coding performance. Also, AVS2 shows competitive performance compared to high efficiency video coding (HEVC). Especially for scene video, AVS2 can achieve 39% bit rate saving over HEVC.

The paper "An Introduction to High Efficiency Video Coding Range Extensions" by Bin Li and Jizheng Xu introduces the coding tools in HEVC range extensions and provides experimental results to compare HEVC range extensions with previous video coding standards.

The paper "Multi-Layer Extension of the High Efficiency Video Coding (HEVC) Standard" by Ming Li *et al.* presents an overview of multi-layer extension of HEVC. With the Multi-layer Extension, HEVC can then have its extension on Scalable HEVC, MV HEVC, 3D HEVC, etc.

The paper "SHVC, the Scalable Extensions of HEVC, and Its Applications" by Yan Ye *et al.* 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.

The paper "ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution" by Peng Yin *et al.* introduces High Dynamic Range (HDR) and Wider Colour Gamut (WCG) content format representation. With advances in display technologies, commercial interest in High Dynamic Range (HDR) and Wide Colour Gamut (WCG) content distribution are growing rapidly. In order to deliver HDR/WCG content, an HDR/WCG video distribution workflow has to be implemented, from content creation to final display.

The paper "DASH and MMT and Their Applications in ATSC 3.0" by Yiling Xu, *et al.* mainly describes features and design considerations of ATSC 3.0 and discusses the applications of the transport protocols used for broadcasting. Additionally, the function of DASH and MMT is to meet the requirement of on-demand viewing of multimedia content over Internet Protocol (IP) with browser-centric media endpoints for more individualized and flexible access to the content.

The paper "Introduction to the AVS2 Scene Video Coding Techniques" by Tiejun Huang *et al.* presents the special applications of AVS2 for surveillance video or

**Guest Editorial**

Huifang Sun, Can Shen, and Ping Wu

video conference videos. By introducing several new coding techniques, AVS2 can provide more efficient compression of scene videos.

The other four papers that have been reviewed and accepted will be published in the next issues due to the page limit.

The paper "Review of AVS Audio Coding Standard" by Tao Zhang *et al.* presents AVS audio coding standard. The latest version of the AVS audio coding standard is ongoing and mainly aims at the increasing demands for low bitrate and high quality audio services. The paper reviews the history and recent development of AVS audio coding standard in terms of basic features, key techniques and performance. Finally, the future development of AVS audio coding standard is discussed.

The paper "Screen Content Coding in HEVC and Beyond" by Tao Lin *et al.* describes an extension of HEVC specially designed to code the videos or pictures captured from a computer screen typically by reading frame buffers or recording digital display output signals of a computer graphics device. Screen content has many unique characteristics not seen in traditional

content. By exploring these unique characteristics, new coding techniques can significantly improve coding performance for screen content.

The paper "Depth Enhancement Methods for Centralized Texture-Depth Packing Formats" by Jar-Ferr Yang *et al.* presents a scheme which can deliver 3D videos through the current 2D broadcasting system with frame-compatible packing formats properly including one texture frame and one depth map in various down-sampling ratios have been proposed to achieve the simplest, most effective solution.

The paper "Light Field Virtual View Rendering based on EPI-representations" by Lu Yu *et al.* presents a new idea for future video coding.

Finally, thank all authors who accepted our invitations and submitted high quality papers in short time with their very busy schedule. Also we take this opportunity to thank all reviewers who provided very valuable comments for further improving the papers. The editors of *ZTE Communications* also made great contributions in this special issue.

**Call for Papers**

*ZTE Communications* Special Issue on  
**Multiple Access Techniques for 5G**

5G mobile cellular networks are required to provide the significant increase in network throughput, cell-edge data rate, massive connectivity, superior spectrum efficiency, high energy efficiency and low latency, compared with the currently deploying long-term evolution (LTE) and LTE-advanced networks. To meet these challenges of 5G networks, innovative technologies on radio air-interface and radio access network (RAN) are important in PHY design. Recently, non-orthogonal multiple access has attracted the interest of both academia and industry as a potential radio access technique.

The upcoming special issue of *ZTE Communications* will focus on the cutting-edge research and application on non-orthogonal multiple access and related signal processing methods for the 5G air-interface. The expected publication date is July 2016. Topics related to this issue include, but are not limited to:

- Non-orthogonal multiple access (NOMA)
- Filter bank multicarrier (FBMC)
- Generalized frequency division multiplexing (GFDM)
- Faster than Nyquist (FTN) transmissions
- Signal detection and estimation in NOMA
- Resource allocations for 5G multiple access
- Cross-layer optimizations of NOMA

- Design and implementation on the transceiver architecture.

**Paper Submission:**

Please directly send to j.yuan@unsw.edu.au and copy to all guest editors, with the subject "ZTE-MAC-Paper-Submission".

**Tentative Schedule:**

Paper submission due: March 31, 2016;  
Review complete: June 15, 2016;  
Final manuscript due: July 31, 2016.

**Guest Editors:**

Prof. Jinhong Yuan, University of New South Wales, Australia (j.yuan@unsw.edu.au)

Dr. Jiying Xiang, ZTE Corporation, China (xiang.jiying@zte.edu.cn)

Prof. Zhiguo Ding, Lancaster University, UK (z.ding@lancaster.ac.uk)

Dr. Liujun Hu, ZTE Corporation, China (hu.liujun@zte.com.cn)

Dr. Zhifeng Yuan, ZTE Corporation, China (yuan.zhifeng@zte.com.cn)



# Overview of the Second Generation AVS Video Coding Standard (AVS2)

Shanshe Wang<sup>1</sup>, Falei Luo<sup>2</sup>, and Siwei Ma<sup>1</sup>

(1. Peking University, Beijing 100871, China;

2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

## Abstract

AVS2 is a new generation video coding standard developed by the AVS working group. Compared with the first generation AVS video coding standard, known as AVS1, AVS2 significantly improves coding performance by using many new coding technologies, e.g., adaptive block partition and two level transform coding. Moreover, for scene video, e.g. surveillance video and conference video, AVS2 provided a background picture modeling scheme to achieve more accurate prediction, which can also make object detection and tracking in surveillance video coding more flexible. Experimental results show that AVS2 is competitive with High Efficiency Video Coding (HEVC) in terms of performance. Especially for scene video, AVS2 can achieve 39% bit rate saving over HEVC.

## Keywords

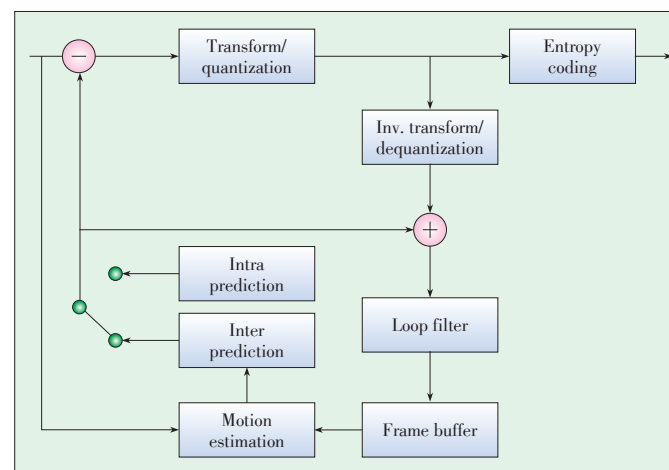
Video Coding; AVS2; AVS1

## 1 Introduction

AVS1 video coding standard, developed by AVS working group, has achieved great success in China and has become an international video coding standard. However, with increased demand for high-resolution videos, it is necessary to develop a new video coding standard that provides much higher coding performance. Based on the success of AVS1 and recent video coding research and standardization, the AVS working group has started the second generation video coding standardization project from 2012, called AVS2. AVS2 is designed to improve coding efficiency for higher resolution videos, and to provide efficient compression solutions for various kinds of video applications, e.g., surveillance video and conference video.

As with previous coding standards, AVS2 uses the traditional hybrid prediction/transform coding framework (Fig. 1). However, AVS2 has more flexible coding tools to satisfy the new requirements identified from emerging applications. First, more flexible prediction block partitions are used to further improve the intra- and inter-prediction accuracy, e.g., square and non-square partitions, which are more adaptive to image content, especially at edge areas. Second, a flexible reference management scheme is proposed to improve inter prediction accuracy. Related to the prediction structure, the transform block size is more flexible and can be up to 64 x 64 pixels. After transforma-

tion, context adaptive arithmetic coding is used for the entropy coding of the transformed coefficients. And a two-level coefficient scan and coding method is adopted to encode the coefficients of large blocks more efficiently. Moreover, for low delay communication applications, e.g., video surveillance, video conferencing, where the background usually does not change often, a background picture model based coding method is developed in AVS2. A background picture constructed from original pictures is used as a reference picture to improve prediction efficiency. Experimental results show that this background



▲ Figure 1. Coding framework of AVS2 encoder.

## Overview of the Second Generation AVS Video Coding Standard (AVS2)

Shanshe Wang, Falei Luo, and Siwei Ma

-picture-based prediction coding can improve the coding efficiency significantly. Furthermore, the background picture can also be used for object detection and tracking for intelligent surveillance video coding.

This paper gives an overview of AVS2 video coding standard and a performance comparison with others. The paper is organized as follows. Section 2 introduces the flexible coding structure in AVS2. Section 3 gives an overview of key tools adopted in AVS2. The specially developed scene video coding is shown in Section 4. Section 5 provides the performance comparison between AVS2 and other state-of-the-art standards. Finally, Section 6 concludes the paper.

## 2 Flexible Coding Structure in AVS2

In AVS2, a flexible coding unit (CU), prediction unit (PU) and transform unit (TU) based coding/prediction/transform structure is used to represent and organize the encoded data [1], [2]. First, pictures are split into largest coding units (LCUs), which consist of  $2N \times 2N$  samples of luminance component and associated chrominance samples with  $N = 8, 16$  or  $32$ . One LCU can be a single CU or can be split into four smaller CUs with a quad-tree partition structure. A CU can be recursively split until it reaches the smallest CU size (**Fig. 2a**). Once the splitting of the CU hierarchical tree is finished, the leaf node CUs can be further split into PUs. A PU is the basic unit for intra- and inter-prediction and allows different shapes to encode irregular image patterns (**Fig. 2b**). The size of a PU is limited to that of a CU with various square or rectangular shapes. Specifically, both intra- and inter-prediction partitions can be symmetric or asymmetric. Intra-prediction partitions

vary in the set  $\{2N \times 2N, N \times N, 2N \times 0.5N, 0.5N \times 2N\}$ , and inter-prediction partitions vary in the set  $\{2N \times 2N, 2N \times N, N \times 2N, 2N \times nU, 2N \times nD, nL \times 2N, nR \times 2N\}$ , where U, D, L and R are the abbreviations of Up, Down, Left and Right respectively.  $n$  is equal to  $0.25N$ . Besides CU and PU, TU is also defined to represent the basic unit for transform coding and quantization. The size of a TU cannot exceed that of a CU, but it is independent of the PU size.

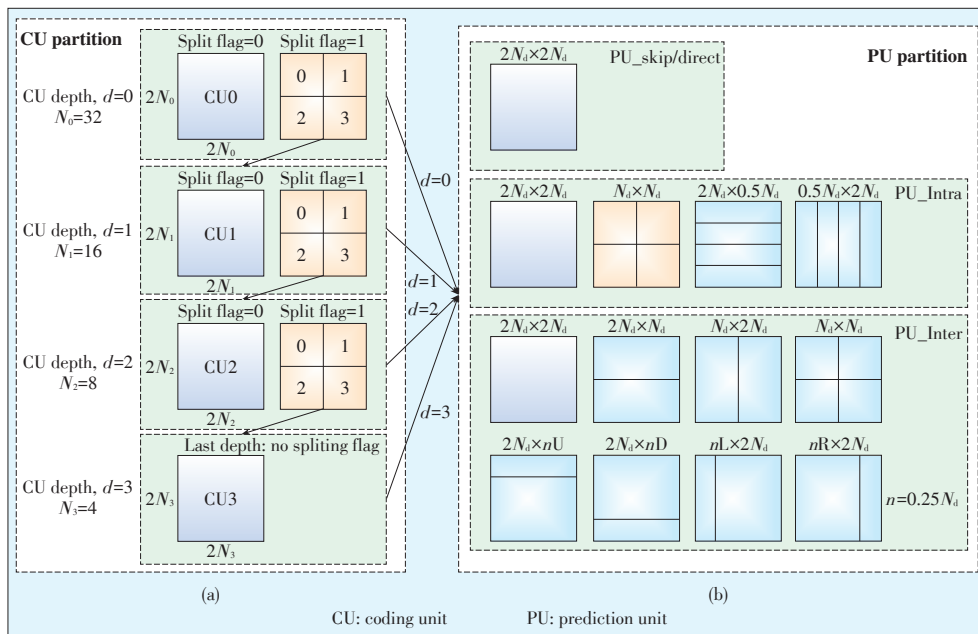
## 3 Main Coding Tools in AVS2

AVS2 uses more efficient coding tools to make full use of the textual information and spatial/temporal redundancies. These tools can be classified into four categories: 1) prediction coding, including intra prediction and inter prediction; 2) transform; 3) entropy coding; and 4) in-loop filtering.

### 3.1 Intra Prediction

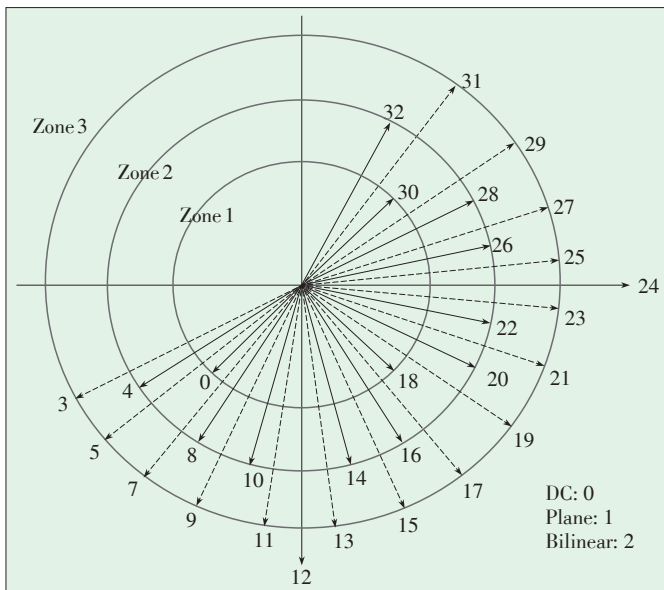
AVS2 still uses a block-partition-based directional prediction to reduce the spatial redundancy in the picture [3]. Compared with AVS1, more intra coding modes are designed to improve the prediction accuracy. Besides the square PU partitions, non-square partitions, called short distance intra prediction (SDIP), are used by AVS2 for more efficient intra luminance prediction [4], where the nearest reconstructed boundary pixels are used as the reference sample in intra prediction (**Fig. 2**). For SDIP, a  $2N \times 2N$  CU is horizontally or vertically partitioned into four PUs. SDIP is more adaptive to the image content, especially in areas with complex textures. To reduce complexity, SDIP is disabled for a  $64 \times 64$  CU. For each prediction block in the partition modes, 33 prediction modes are supported for luminance, including

30 angular modes [3], plane mode, bilinear mode and DC mode. As in **Fig. 3**, the prediction directions associated with the 30 angular modes are distributed within the range of  $[-157.5^\circ, 60^\circ]$ . Each sample in a PU is predicted by projecting its location to the reference pixels in the selected prediction direction. To improve intra-prediction accuracy, the sub-pixel precision reference samples are interpolated if the projected reference samples locate on a non-integer position. The non-integer position is bounded to  $1/32$  sample precision to avoid floating point operation, and a 4-tap linear interpolation filter is used to obtain the sub-pixel. During the coding of luma prediction mode, two most probable modes (MPMs) are used for



▲ **Figure 2.** (a) Maximum possible recursive CU structure in AVS2 (LCU size= 64, maximum hierarchical depth = 4), (b) Possible PU splitting for skip, intra and inter modes in AVS2.





▲ Figure 3. Illustration of directional prediction modes.

prediction. If the current prediction mode equals one of the MPMs, two bins are transmitted into the bitstream; otherwise, six bins are needed.

For the chrominance component, the PU is always square, and 5 prediction modes are supported, including vertical prediction, horizontal prediction, bilinear prediction, DC prediction and the prediction mode derived from the corresponding luminance prediction mode [5].

### 3.2 Inter Prediction

Compared to the spatial intra prediction, inter prediction focuses on exploiting the temporal correlation between the consecutive pictures to reduce the temporal redundancy. AVS2 still adopts the multi-reference prediction as in AVS1, including both short term and long term reference pictures. However, inter prediction mode has been improved much and a more flexible reference picture management scheme is adopted.

#### 3.2.1 Improved Inter-Prediction Mode

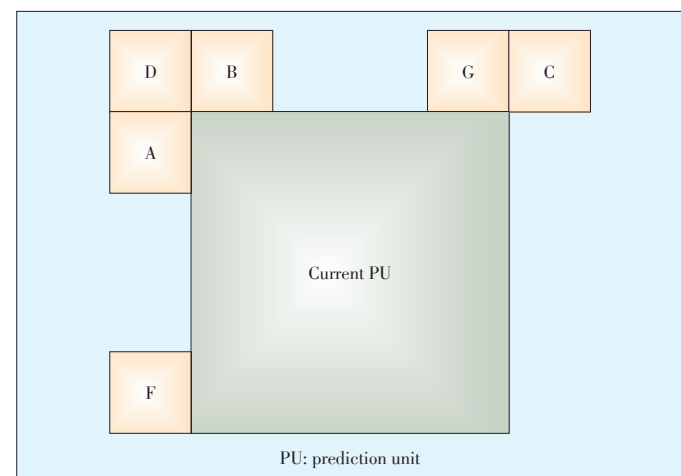
In AVS2, inter prediction mode has been improved much to further improve the inter prediction accuracy. Firstly, a new inter frame type, called F frame, is defined as a special P frame [6] in addition to the traditional P and B frames. Secondly, new inter coding modes are specially designed for F and B frame.

For F frame, besides the conventional single hypothesis prediction mode as in a P frame, the significant improvement is the use of multi-hypothesis techniques, including multi-directional skip/direct mode [7], temporal multi-hypothesis prediction mode [8], and spatial directional multi-hypothesis (DMH) prediction mode [9]. These modes improve the coding performance of AVS2 by a large margin. Detailed descriptions are shown as follows.

The multi-directional skip/direct mode in F frame is used to

merge current block to spatial or temporal neighboring block. The difference between skip mode and direct mode is that skip mode needs to encode residual information while direct mode does not. However, the derivation of motion vector (MV) for the two modes are the same. In AVS2, two derivation methods, one of which is temporal and the other is spatial, are used. For temporal derivation, one MV is achieved from the temporal collocated block in the nearest or backward reference frame. The other MV for weighted skip mode is obtained by scaling the first derived MV in the second reference frame. The second reference is specified by the reference index transmitted in the bitstream, indicating weighted skip mode. For spatial derivation, the needed motion vectors, one or two, are obtained from neighboring prediction blocks. If only one MV is needed, two derivations are provided. One is to search the neighboring blocks (Fig. 4) in a pre-defined order: F, G, C, A, B, D. The other is to determine the MV by searching the neighboring blocks in a reverse order. If the derived MVs do not belong to the same block, the two MVs are available. Otherwise, the second MV should be re-derived from the neighboring blocks using dual forward prediction. If two MVs are needed, the derivation scheme is the same as before. The difference is that when the two MVs belong to the same block, the second MV should re-derive by combining one MV single forward prediction searched by the defined order and one MV searched by reversed order.

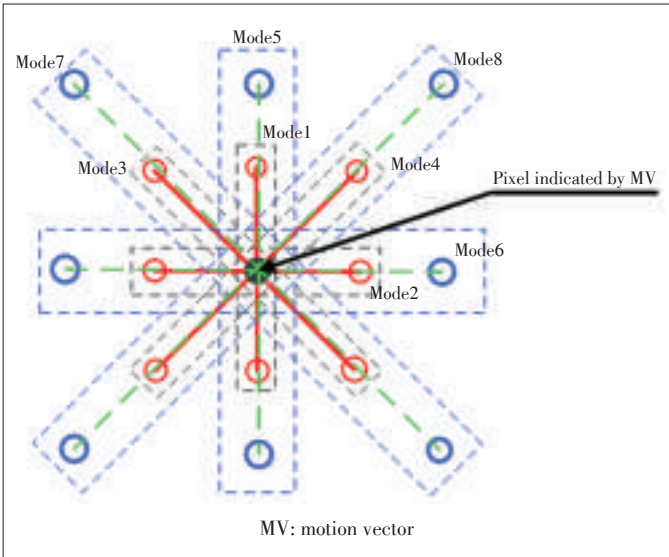
DMH mode provides a derivation scheme to generate two seed predictors based on the initial predictor obtained from motion estimation to improve the inter prediction accuracy. As in Fig. 5, all the optional seed predictors are located on the line crossing the initial predictor. Considering the coding complexity, the number of seed predictors is restricted to 8, mode 1 to mode 8. The derivation of the two seed predictors is shown in Table 1. For one seed predictor mode with index as  $i$ , MV offset, denoted as  $\overline{dmh}_i$ , is firstly obtained according to the table. Then the needed two seed predictors,  $\overline{mv}_1$  and  $\overline{mv}_2$ , are calc-



▲ Figure 4. Illustration of neighboring blocks A, B, C, D, F and G for motion vector prediction.

Overview of the Second Generation AVS Video Coding Standard (AVS2)

Shanshe Wang, Falei Luo, and Siwei Ma



▲ Figure 5. DMH mode.

▼ Table 1. The derivation of seed predictors for DMH

Mode index	$\overrightarrow{dmh}_i$
1	(1, 0)
2	(0, 1)
3	(1, -1)
4	(1, 1)
5	(2, 0)
6	(0, 2)
7	(2, -2)
8	(2, 2)

ulated based on the original ( $\overrightarrow{mv}_o$ ) as follows.

$$\overrightarrow{mv}_1 = \overrightarrow{mv}_o + \overrightarrow{dmh}_i$$

(1)

$$\overrightarrow{mv}_2 = \overrightarrow{mv}_o - \overrightarrow{dmh}_i$$

(2)

For B frame, the coding modes are also expanded to improve prediction accuracy. In addition to the conventional forward, backward, bi - directional and skip/direct prediction modes, symmetric prediction is defined as a special bi - prediction mode, wherein only one forward-motion vector is coded, and the backward motion vector is derived from the forward motion vector.

3.2.2 Flexible Reference Picture Management

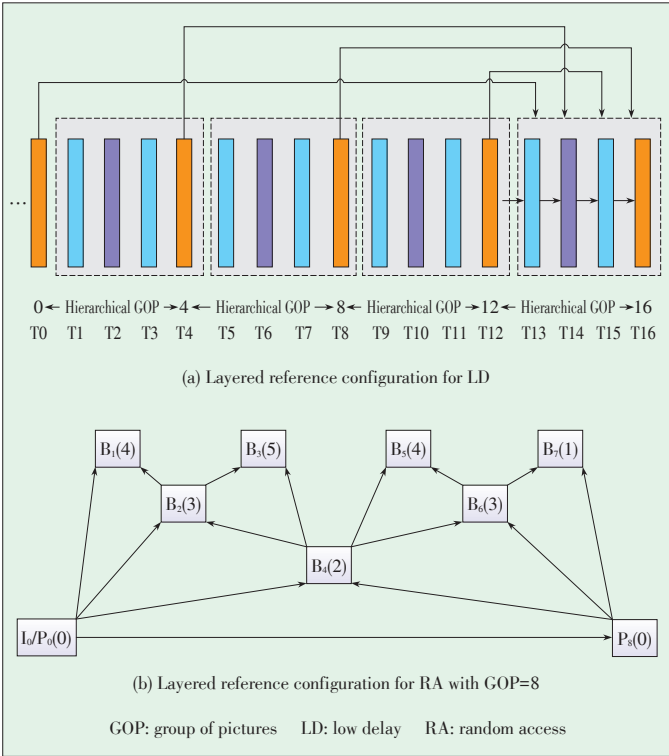
AVS2 adopts a flexible reference picture management scheme to improve the inter prediction efficiency. In the scheme, a reference configuration set (RCS) is used to manage the reference pictures. RCS consists of reference picture information of current coding picture, including decoding order index (DOI), QP offset, number of reference pictures, delta DOIs between current picture and reference pictures, number of pic-

tures that need to remove from buffer and delta DOIs between pictures to remove and current pictures.

In order to save coding bits, several RCS sets are used and signaled in the sequence header. Only the index of RCS is transmitted in the picture header. Based on RCS, the reference picture set for current coding picture can be arbitrarily configured. Fig. 6 shows the layered reference configuration on AVS2.

3.3 Motion Vector Prediction and Coding

The motion vector prediction (MVP) plays an important role in inter prediction, which can reduce redundancy between motion vectors of neighbor blocks and save many coding bits for motion vectors. In AVS2, four different prediction methods are adopted (Table 2). Each of these has its unique usage. Spatial motion vector prediction is used for spatial derivation of Skip/



Direct mode in F frames and B frames. Temporal motion vector prediction is used for temporal derivation of Skip/Direct mode in all inter frames. Spatial-temporal combined motion vector prediction is used for temporal derivation of Skip/Direct mode in B frames. For other cases, median prediction is used. Moreover, in order to improve the MV prediction accuracy, the derivation of MV is achieved by the reference distance based scaling.

In AVS2, the motion vector is in quarter-pixel precision for the luminance component, and the sub-pixel is interpolated with an 8-tap DCT interpolation filter (DCT-IF) [10]. For the chrominance component, the motion vector derived from luminance with 1/8 pixel precision and a 4-tap DCT-IF is used for sub-pixel interpolation [11]. The filter coefficients for sub-pixel interpolation is defined in **Table 3**. After motion vector prediction, the motion vector difference (MVD) is coded in the bit-stream. However, redundancy may still exist in MVD, and to further save coding bits of motion vectors, a progressive motion vector resolution (PMVR) adaptation method is used in AVS2 [12]. In PMVR, MVP is first rounded to the nearest half sample position, and then the MVD is rounded to half-pixel precision if it exceeds a threshold. Furthermore, the resolution of MVD is decreased to integer-pel precision if it exceeds another threshold. In AVS2, only one threshold is used, which means that if the distance between the MV and MVP is less than the threshold, quarter-pixel based MVD is coded; otherwise, half-pixel based MVD is coded (actually, the MVD is separated into two parts and coded with different resolution. The part of MVD within the window will be coded at 1/4 pixel resolution, and the other part will be coded at half-pixel resolution).

### 3.4 Transform

Unlike the transform in AVS1, a flexible TU partition structure is used to further compress the predicted residual in AVS2. For CU with symmetric prediction unit partition, the TU size can be  $2N \times 2N$  or  $N \times N$  signaled by a transform split flag. For CU with asymmetric prediction unit partition, the TU size can be  $2N \times 2N$ ,  $n \times 2N$  or  $2N \times n$ . Thus, the maximum transform

▼ **Table 3.** DCT-like interpolation filter for sub-pixel interpolation

Interpolation	Position	Coefficients
Luma	1/4	{ -1, 4, -10, 58, 17, -5, 1, 0 }
	2/4	{ -1, 4, -11, 40, 40, -11, 4, -1 }
	3/4	{ 0, 1, -5, 17, 58, -10, 4, -1 }
Chroma	1/8	{ -4, 62, 6, 0 }
	2/8	{ -6, 56, 15, -1 }
	3/8	{ -5, 47, 25, -3 }
	4/8	{ -4, 36, 36, -4 }
	5/8	{ -3, 25, 47, -5 }
	6/8	{ -1, 45, 56, -6 }
	7/8	{ 0, 6, 62, -4 }

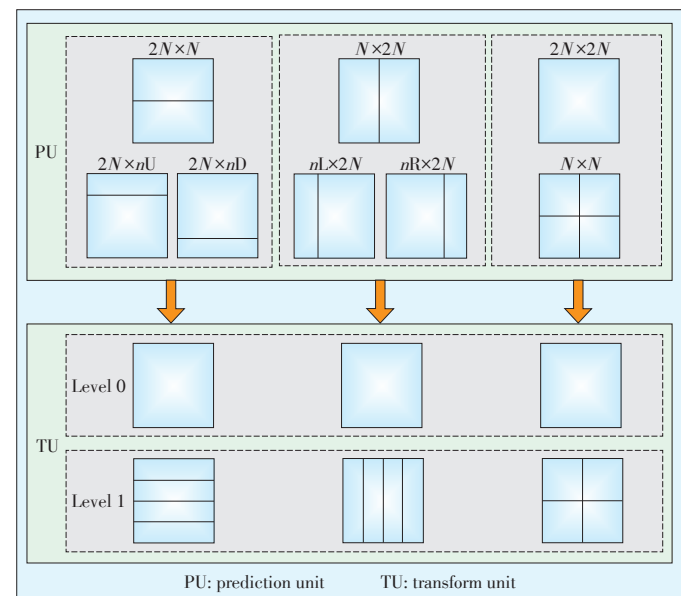
size is  $64 \times 64$ , and the minimum is  $4 \times 4$ . For TU size from  $4 \times 4$  to  $32 \times 32$ , an integer transform (IT) that closely approximates the performance of the discrete cosine transform (DCT) is used. For a square residual block, the forward transform matrices from  $4 \times 4$  to  $32 \times 32$ . Here,  $4 \times 4$  transform  $T_4$  and  $8 \times 8$  transform  $T_8$  are:

$$T_4 = \begin{bmatrix} 32 & 32 & 32 & 32 \\ 42 & 17 & -17 & 42 \\ 32 & -32 & -32 & 32 \\ 17 & -42 & 42 & -17 \end{bmatrix} \quad (3)$$

$$T_8 = \begin{bmatrix} 32 & 32 & 32 & 32 & 32 & 32 & 32 & 32 \\ 44 & 38 & 25 & 9 & -9 & -25 & -38 & -44 \\ 42 & 17 & -17 & -42 & -42 & -17 & 17 & 42 \\ 38 & -9 & -44 & -25 & 25 & 44 & 9 & -38 \\ 32 & -32 & -32 & 32 & 32 & -32 & -32 & 32 \\ 25 & -44 & 9 & 38 & -38 & -9 & 44 & -25 \\ 17 & -42 & 42 & -17 & -17 & 42 & -42 & 17 \\ 9 & -25 & 38 & -44 & 44 & -38 & 25 & -9 \end{bmatrix} \quad (4)$$

For a  $64 \times 64$  transform, a logical transform (LOT) [13] is applied to the residual. A 5-3 tap integer wavelet transform is first performed on a  $64 \times 64$  block discarding the LH, HL and HH-bands, and then a normal  $32 \times 32$  IT is applied to the LL-band. For all the PU partitions of a CU,  $2N \times 2N$  IT is used in the first level, and a non-square transform [14] is used in the second level (**Fig. 7**).

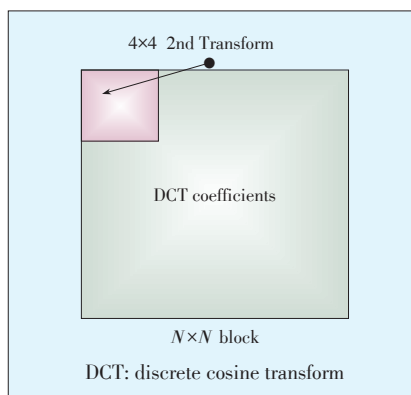
Furthermore, a secondary transform can be used to reduce the correlation for luminance intra-prediction residual block. The secondary transform matrix is related to the block size. If the transform block size is greater than or equal to  $8 \times 8$ , a  $4 \times 4$  secondary transform with matrix  $S_4$  is applied to the left corner of the transform block as shown in **Fig. 8**. If the transform block size is  $4 \times 4$ , an independent transform matrix  $D_4$  rather



▲ **Figure 7.** PU partition and two-level transform coding.

## Overview of the Second Generation AVS Video Coding Standard (AVS2)

Shanshe Wang, Falei Luo, and Siwei Ma



◀ Figure 8. Illustration of secondary transform in AVS2.

than  $T_4$  is used.

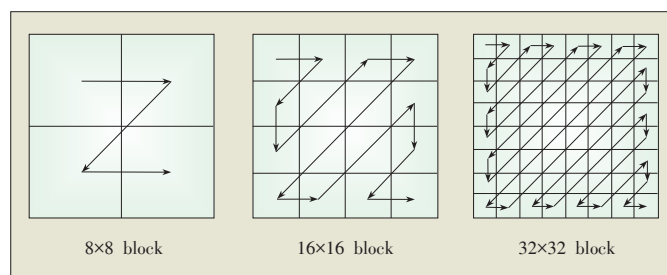
$$S_4 = \begin{bmatrix} 123 & -35 & -8 & -3 \\ -32 & -120 & 30 & 10 \\ 14 & 25 & 123 & -22 \\ 8 & 13 & 19 & 126 \end{bmatrix}, D_4 = \begin{bmatrix} 34 & 58 & 72 & 81 \\ 77 & 69 & -7 & -75 \\ 79 & -33 & -75 & 58 \\ 55 & -84 & 73 & -28 \end{bmatrix} \quad (5)$$

### 3.5 Entropy Coding

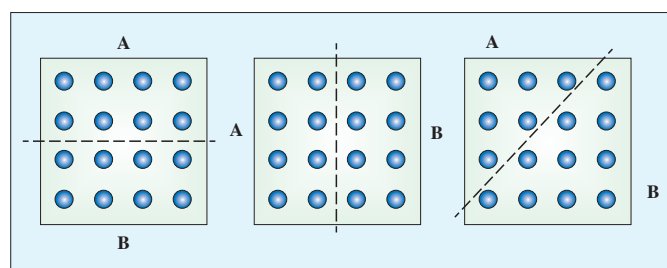
The entropy coding used in AVS2 is inherited from AVS1. The arithmetic coding engine is designed according to a logarithmic model. Thus, the probability estimation is specified to be multiplication-free and only using shifts and addition and no look-up tables are needed.

For the transformed coefficients coding, a two-level coding scheme is applied to the transform coefficient blocks [15]. First, a coefficient block is partitioned into 4x4 coefficient groups (CGs) (Fig. 9). Then zig-zag scanning and Context-Based Adaptive Binary Arithmetic Coding (CABAC) is performed at both the CG level and coefficient level. At the CG level for a TU, the CGs are scanned in zig-zag order, and the CG position indicating the position of the last non-zero CG is coded first, followed by a bin string in the reverse zig-zag scan order of significant CG flags indicating whether the CG contains non-zero coefficients. At the coefficient level, for each non-zero CG, the coefficients are further scanned into the form of (run, level) pair in zig-zag order. Level and run indicate the magnitude of a non-zero coefficient and the number of zero coefficients between two non-zero coefficients, respectively. For the last CG, the coefficient position, which denotes the position of the last non-zero coefficient in scan order, is coded first. For a non-last CG, a last run is coded which denotes number of zero coefficients after the last non-zero coefficient in zig-zag scan order. Then the (level, run) pairs in a CG are coded in reverse zig-zag scan order.

For the context modeling, AVS2 uses a mode-dependent context-selection design for intra-prediction blocks [16]. In this context design, 33 intra-prediction modes are classified into three prediction mode sets: vertical, horizontal, and diagonal. Depending on the prediction mode set, each CG is divided to two regions (Fig. 10). The intra-prediction modes and CG re-



▲ Figure 9. Sub-block scan for transform blocks of size 8x8, 16x16 and 32x32 transform blocks; each sub-block represents a 4x4 coefficient group.



▲ Figure 10. Sub-block region partitions of 4x4 coefficient group in an intra prediction block.

gions are applied in the context modeling of syntax elements including the last CG position, last coefficient position and run value. In addition, AVS2 takes more consideration on data dependence reduction in context design and explores more possibility for bypass mode as well.

### 3.6 In-Loop Filtering

Compared to AVS1, AVS2 has made great improvement over in-loop filtering. Except for de-blocking filter, two more filtering processes are added to AVS2, called sample adaptive offset (SAO) filtering [17] and adaptive loop filter (ALF) [18], to further improve the reconstructed picture quality. Thus in-loop filtering in AVS2 includes the following three sequential procedures: deblocking filtering, SAO and ALF.

The deblocking filter is designed to remove the blocking artifacts caused by block transform and quantization. In AVS2, the basic unit for deblocking filter is an 8x8 block. For each 8x8 block, deblocking filter is used only if the boundary belongs to either of CU boundary, PU boundary or TU boundary. Unlike AVS1, gradient is considered for boundary strength (BS) calculation and then BS is classified into more levels based on the calculated gradient. When the boundary is not the edge of a block which can be CU, PU or TU, BS is set to the lowest value to reduce the complexity.

After the deblocking filter, an SAO filter is applied to reduce the mean sample distortion of a region. The basic unit of SAO is defined as four pixels top-left the LCU region, which is more flexible for parallelization. An offset is added to the reconstructed sample for each SAO filter unit to reduce ringing artifacts and contouring artifacts. There are two kinds of offset

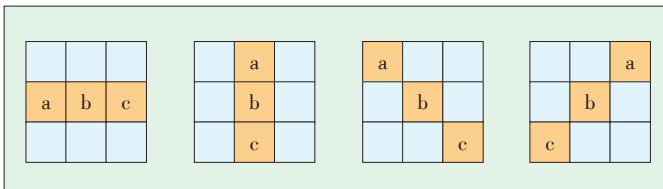
called Edge Offset (EO) and Band Offset (BO) mode, respectively.

Edge Offset mode first classifies the pixels in the filter unit using four 1-D directional patterns as illustrated in **Fig. 11**. According to these patterns, four EO classes are specified, and only one EO class can be selected for each filter unit. For a given EO class, samples of current filter unit are classified into one of the five categories, which are based on the rules defined in **Table 4**. For pixels in each category except category 0, an offset is derived by calculating the mean of the difference of reconstructed pixel values and original pixel values. The offset and the index of classification pattern are transmitted to a decoder.

Band offset mode classifies the pixels into 32 bands by equally dividing the pixel range. Theoretically, one offset can be derived for each band by calculating the mean of the difference of reconstructed pixel values and original pixel values. However, more coding bits are necessary. Statistical results show that the offsets of most pixel belong to a small domain. Thus in AVS2, only four bands are selected in order to save coding bits. Considering the fact that some sample values may be quite different with the others, 2 start band positions are transmitted to the decoder.

Besides EO and BO, merge technique is utilized in order to save the bits consuming, where a merge flag is employed to indicate whether the SAO parameters of the current LCU is exact the same with its neighbors. When merge flag is enabled, all the following SAO parameters are not signaled but inferred from neighbors.

ALF is the last stage of in-loop filtering. Its nature is to minimize the mean squared error between the original frame and the reconstructed frame using Wiener-Hopf equations. There are two stages in this process at encoder side. The first stage is filter coefficient derivation. To achieve the filter coefficients, reconstructed pixels of the luminance component are classified into 16 categories, and one set of filter coefficients is trained



▲ Figure 11. Four 1-D directional EO patterns.

▼ Table 4. The classification rules and pixel categories

Category	Condition	Offset Range
1	$c < a \ \&\& \ c < b$	$-1 \leq \text{offset} \leq 6$
2	$(c < a \ \&\& \ c == b) \parallel (c == a \ \&\& \ c < b)$	$0 \leq \text{offset} \leq 1$
3	$(c > a \ \&\& \ c == b) \parallel (c == a \ \&\& \ c > b)$	$-1 \leq \text{offset} \leq 0$
4	$c > a \ \&\& \ c > b$	$-6 \leq \text{offset} \leq 1$
0	None of the above	None

for each category using Wiener-Hopf equations. To reduce the redundancy between these 16 sets of filter coefficients, the encoder will adaptively merge them based on the rate-distortion performance. At its maximum, 16 different filter sets can be assigned for the luminance component and only one for each chrominance component. The second stage is to filter each sample with the corresponding derived filter coefficients using a 7x7 cross and 3x3 square filter as shown in **Fig. 12**.

Finally, the filtered sample can be achieved as follows:

$$ptmp = C[\text{filterIdx}][8] \times p(x,y) + \sum_{j=0}^7 C[\text{filterIdx}][j] \times (p(x - \text{Hor}[j], y - \text{Ver}[j]) + p(x + \text{Hor}[j], y + \text{Ver}[j])) \quad (6)$$

$$ptmp = (ptmp + 32) \gg 6 \quad (7)$$

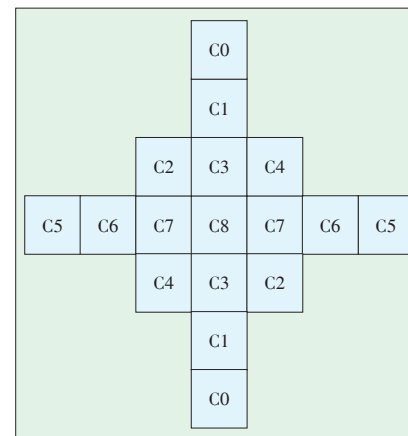
$$p'(x,y) = \text{Clip3}(0, (1 \ll \text{BitDepth}) - 1, ptmp) \quad (8)$$

where  $\text{filterIdx}$  indicates luma or chroma component,  $p(x,y)$  is the reconstructed sample after SAO.  $p'(x,y)$  is the final reconstructed sample after ALF.  $\text{Hor}[j]$  and  $\text{Ver}[j]$  stands for the filter coefficients positions.

## 4 Scene Video Coding

In practical applications, many videos are captured in specific scenes, such as surveillance video and videos from classroom, home, court, etc., which are characterized by temporally stable background. The redundancy originating from the background could be further reduced. In AVS2, a background-picture-model-based coding method is proposed to achieve higher compression performance [19] (**Fig. 13**). G-pictures and S-pictures are defined to further exploit the temporal redundancy and facilitate video event generation such as object segmentation and motion detection.

The G-picture is a special I-picture, which is stored in a separate background memory. It is encoded by intra mode only and is not decoded for displaying. The reason is that it is just for being referenced rather than for viewing. For the generation of a G-picture, a method of segment-and-weight based running

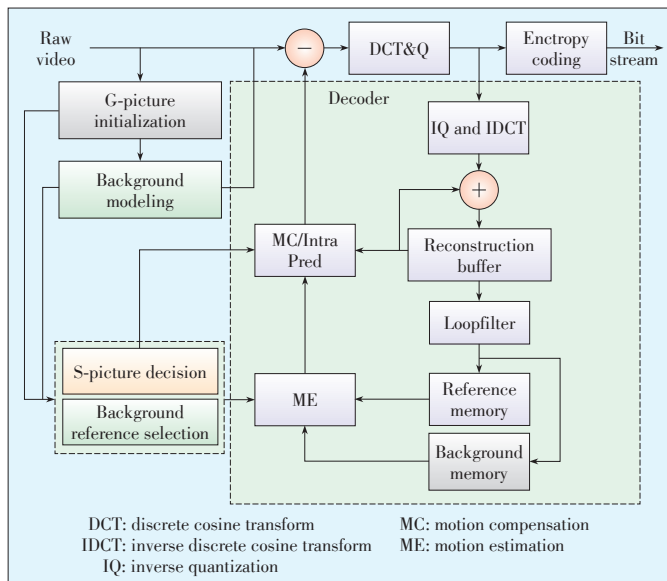


◀ Figure 12. Adaptive loop filter shape.



## Overview of the Second Generation AVS Video Coding Standard (AVS2)

Shanshe Wang, Falei Luo, and Siwei Ma



▲ Figure 13. Background picture based scene coding in AVS2.

average (SWRA) [20] is used to generate the GB picture. SWRA approximately generates the background by assigning larger weights on the frequent values in the averaging process. When encoding the G-picture, a smaller QP is selected to make a high-quality G-picture. Then the G-picture can be well referenced by the following pictures.

S-picture is a special P-picture that can only be predicted from a reconstructed G-picture or virtual G-picture which does not exist in the actual input sequence but is modeled from input pictures and encoded into the stream to act as a reference picture. Only intra, SKIP and P2N×2N modes with zero motion vectors are available in S picture. In the AVS2, the S picture is set as the random access point instead of intra-predicted picture. However, the S picture outperforms I picture when adopted as the random access point since the inter prediction is adopted in the S picture and the prediction performance is better. With the S-picture, the performance of the random access can be improved on a large scale.

Furthermore, according to the predication modes in AVS2 compression bitstream, the blocks of an AVS2 picture could be classified as background blocks, foreground blocks or blocks on the edge area. Obviously, this information is very helpful for possible subsequent vision tasks, such as object detection and tracking. Object-based coding has already been proposed in MPEG-4; however, object segmentation remains a challenge that constrains the application of object based coding. Therefore, AVS2 uses simple background modeling instead of accurate object segmentation. The simple background modeling is easier and provides a good tradeoff between coding efficiency and complexity.

To provide convenience for applications like event detection and searching, AVS2 adds some novel high-level syntax to describe the region of interest (ROI). In the region extension, the

region number, event ID, and coordinates for top left and bottom right corners are included to show what number the ROI is, what event happened and where it lies.

## 5 Performance Comparison

In this section, the performance comparisons among AVS2, AVS1, and state-of-the-art High Efficiency Video Coding (HEVC) international standard are provided. For comparison, the reference software used in the experiments is HM16.6 for HEVC, GDM 4.1 for AVS1 and RD12.0 for AVS2. HEVC and AVS1 are used as a testing anchor. According to the applications, we tested the performance of AVS2 with three different coding configurations: all-intra (AI), random access (RA), and low delay (LD), similar to the HEVC common test conditions and BD-Rate is used for bitrate saving evaluation. The UHD, 1080 p, 720 p, WVGA and WQVGA test sequences are the common test sequences used in AVS2, including partial test sequences used in HEVC, such as Traffic (UHD), Kimono1 (1080 p), BasketballPass (WQVGA) and City (720 p). Moreover, surveillance sequences including 1200 p and 576 p are tested to further compare the performance of AVS2 and HEVC under their respective common test condition. All these sequences and the surveillance/videoconference sequences are available on the AVS website.

Table 5 shows the rate distortion performance of AVS2 for three test cases. For different test configurations, AVS2 shows comparable performance as HEVC and outperforms AVS1 with significant bits saving, up to 52.9% for RA. Table 6 shows the rate distortion performance comparisons of AVS2 with HEVC for surveillance sequences. AVS2 outperforms HEVC by

▼ Table 5. Bitrate saving of AVS2 performance comparison with AVS1, HEVC for common test sequences

Sequences	AI configuration		RA configuration		LD configuration	
	AVS1 vs. AVS2	HEVC vs. AVS2	AVS1 vs. AVS2	HEVC vs. AVS2	AVS1 vs. AVS2	HEVC vs. AVS2
UHD	-31.2%	-2.21%	-50.5%	-0.29%	-57.6%	2.72%
1080 p	-33.1%	-0.67%	-51.3%	-2.30%	-44.3%	0.68%
720 p	-34.0%	-2.06%	-57.2%	-2.44%	-56.3%	1.88%
WVGA	-30.4%	1.46%	-52.8%	0.05%	-50.5%	0.91%
WQVGA	-26.6%	2.78%	-52.4%	1.08%	-49.4%	4.87%
Overall	-31.2%	-0.06%	-52.9%	-0.88%	-51.0%	2.11%

HEVC: High Efficiency Video Coding AI: all-intra LD: low delay RA: random access

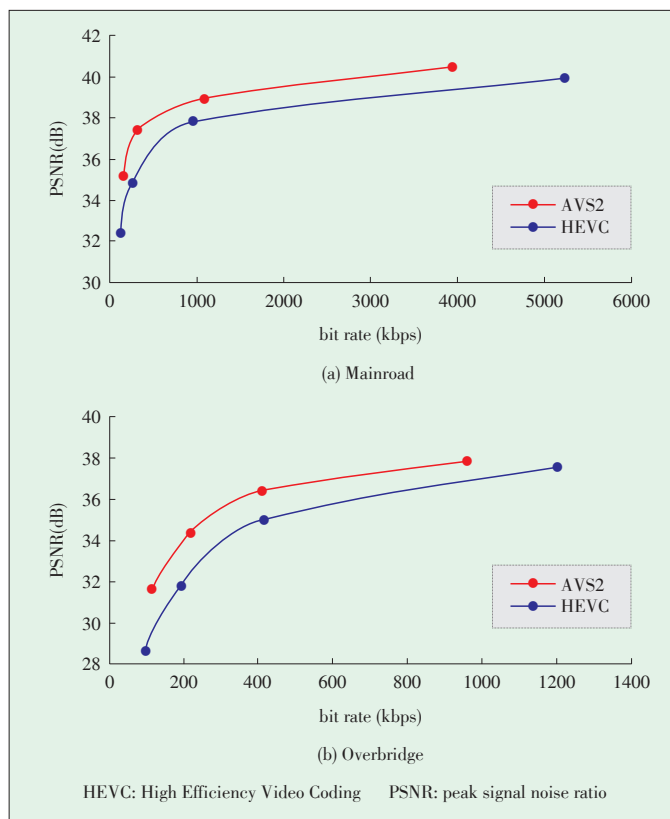
▼ Table 6. Bitrate saving of AVS2 performance comparison with HEVC for surveillance sequences

Sequences	RA configuration	LD configuration
1200 p	-35.7%	-38.5%
576 p	-41.3%	-26.5%
Overall	-39.1%	-31.3%

LD: low delay RA: random access

## Overview of the Second Generation AVS Video Coding Standard (AVS2)

Shanshe Wang, Falei Luo, and Siwei Ma



▲ Figure 14. Performance comparison between AVS2 and HEVC for surveillance videos: (a) MainRoad, (b) Overbridge.

39.1% and 31.3% under RA and LD test configuration, respectively. The curves in Fig. 14 show the results on two surveillance video sequences.

## 6 Conclusions

This paper gives an overview of the AVS2 standard. AVS2 is an application oriented coding standard, and different coding tools have been developed according to various application characteristics and requirements. For high quality broadcasting, flexible prediction and transform coding tools have been incorporated. Especially for scene applications, AVS2 significantly improves coding performance and bridges video compression with machine vision by incorporating the background picture modeling, thereby making video coding smarter and more efficient. In a word, compared to the previous AVS1 coding standard, AVS2 achieves significant improvement both in coding efficiency and flexibility.

## References

- [1] S. Ma, S. Wang, and W. Gao, "Overview of IEEE 1857 video coding standard," in *Proc. IEEE International Conference on Image Processing*, Melbourne, Australia, Sept. 2013, pp.1500–1504. doi: 10.1109/MSP.2014.2371951.
- [2] Q. Yu, S. Ma, Z. He, et al., "Suggested video platform for AVS2," 42nd AVS Meeting, Guilin, China, AVS\_M2972, Sept. 2012.

- [3] Y. Piao, S. Lee and C. Kim, "Modified intra mode coding and angle adjustment," 48th AVS Meeting, Beijing, China, AVS\_M3304, Apr. 2014.
- [4] Q. Yu, X. Cao, W. Li, et al., "Short distance intra prediction," 46th AVS Meeting, Shenyang, China, AVS\_M3171, Sept. 2013.
- [5] Y. Piao, S. Lee, I.-K. Kim, and C. Kim, "Derived mode (DM) for chroma intra prediction," 44th AVS Meeting, Luoyang, China, AVS\_M3042, Mar. 2013.
- [6] Y. Lin and L. Yu, "F frame CE: Multi forward hypothesis prediction," 48th AVS Meeting, Beijing, China, AVS\_M3326, Apr. 2014.
- [7] Z. Shao and L. Yu, "Multi-hypothesis skip/direct mode in P frame," 47th AVS Meeting, Shenzhen, China, AVS\_M3256, Dec. 2013.
- [8] Y. Ling, X. Zhu, L. Yu, et al., "Multi-hypothesis mode for AVS2," 47th AVS meeting, Shenzhen, China, AVS\_M3271, Dec. 2013.
- [9] I.-K. Kim, S. Lee, Y. Piao, and C. Kim, "Directional multi-hypothesis prediction (DMH) for AVS2," 45th AVS Meeting, Taicang, China, AVS\_M3094, Jun. 2013.
- [10] H. Lv, R. Wang, Z. Wang, et al., "Sequence level adaptive interpolation filter for motion compensation," 47th AVS Meeting, Shenzhen, China, AVS\_M3253, Dec. 2013.
- [11] Z. Wang, H. Lv, X. Li, et al., "Interpolation improvement for chroma motion compensation," 48th AVS Meeting, Beijing, China, AVS\_M3348, Apr. 2014.
- [12] J. Ma, S. Ma, J. An, K. Zhang, and S. Lei, "Progressive motion vector precision," 44th AVS Meeting, Luoyang, China, AVS\_M3049, Mar. 2013.
- [13] S. Lee, I.-K. Kim, Min-Su Cheon, N. Shlyakhov, and Y. Piao, "Proposal for AVS2.0 Reference Software," 42nd AVS Meeting, Guilin, China, AVS\_M2973, Sept. 2012.
- [14] W. Li, Y. Yuan, X. Cao, et al., "Non-square quad-tree transform," 45th AVS Meeting, Taicang, China, AVS\_M3153, Jun. 2013.
- [15] J. Wang, X. Wang, T. Ji, and D. He, "Two-level transform coefficient coding," 43rd AVS Meeting, Beijing, China, AVS\_M3035, Dec. 2012.
- [16] X. Wang, J. Wang, T. Ji, and D. He, "Intra prediction mode based context design," 45th AVS Meeting, Taicang, China, AVS\_M3103, Jun. 2013.
- [17] J. Chen, S. Lee, C. Kim, et al., "Sample adaptive offset for AVS2," 46th AVS Meeting, Shenyang, China, AVS\_M3197, Sept. 2013.
- [18] X. Zhang, J. Si, S. Wang, et al., "Adaptive loop filter for AVS2," 48th AVS Meeting, Beijing, China, AVS\_M3292, Apr. 2014.
- [19] S. Dong, L. Zhao, P. Xing, and X. Zhang, "Surveillance video coding platform for AVS2," 47th AVS Meeting, Shenzhen, China, AVS\_M3221, Dec. 2013.
- [20] X. Zhang, Y. Tian, T. Huang, and W. Gao, "Low-complexity and high efficiency background modelling for surveillance video coding," in *IEEE International Conference on Visual Communication and Image Processing*, San Diego, USA, Nov. 2012, pp. 1–6. doi: 10.1109/VCIP.2012.6410796.

Manuscript received: 2015-11-16

## Biographies

**Shanshe Wang** (sswang@pku.edu.cn) received the BS degree in Department of Mathematics from Heilongjiang University, China in 2004, MS degree in computer software and theory from Northeast Petroleum University, China in 2010, and PhD degree in computer science from the Harbin Institute of Technology, China. Now he is a post doctor of Computer Science, National Engineering Lab. on Video Technology, Peking University, China. His current research interests include video compression and image and video quality assessment.

**Falei Luo** (falei.luo@vip.163.com) received the BS degree from Huazhong University of Science and Technology, China and is currently pursuing the PhD degree at Institute of Computing Technology, Chinese Academy of Sciences, China.

**Siwei Ma** (swma@pku.edu.cn) received the BS degree from Shandong Normal University, China in 1999, and the PhD degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, China in 2005. From 2005 to 2007, he held a post-doctorate position with the University of Southern California, Los Angeles, USA. Then, he joined the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, China, where he is currently a professor of Computer Science, National Engineering Lab. on Video Technology, and a co-chair of AVS video Subgroup. He has published over 100 technical articles in refereed journals and proceedings in the areas of image and video coding, video processing, video streaming, and transmission.

# An Introduction to High Efficiency Video Coding Range Extensions

**Bin Li and Jizheng Xu**

(Microsoft Research Asia, Beijing 100080, China)

## Abstract

High Efficiency Video Coding (HEVC) is the latest international video coding standard, which can provide the similar quality with about half bandwidth compared with its predecessor, H.264/MPEG-4 AVC. To meet the requirement of higher bit depth coding and more chroma sampling formats, range extensions of HEVC were developed. This paper introduces the coding tools in HEVC range extensions and provides experimental results to compare HEVC range extensions with previous video coding standards. Experimental results show that HEVC range extensions improve coding efficiency much over H.264/MPEG-4 AVC High Predictive profile, especially for 4K sequences.

## Keywords

H.265; High Efficiency Video Coding (HEVC); MPEG-H; range extensions; video compression

## 1 Introduction

**H**igh Efficiency Video Coding (HEVC) [1] is the latest international video coding standard, standardized as ITU-T Recommendation H.265 and ISO/IEC 23008-1 (MPEG-H Part 2). Compared with its predecessor, H.264/MPEG-4 Advanced Video Coding (AVC) [2], about 50% bit saving can be achieved [3]. Although HEVC version 1 supports a wide variety of applications, some key features are not included and left for further developments.

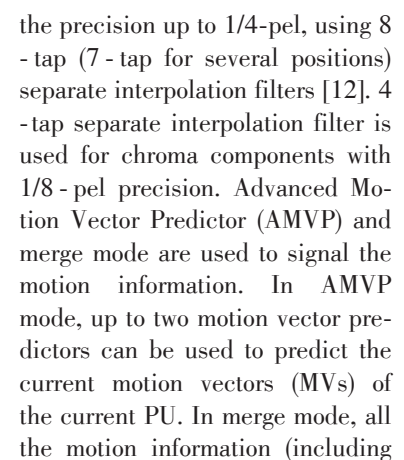
After the finalization of HEVC version 1, several extensions of HEVC are being developed. Of these, Range Extensions (RExt) support various chroma sampling formats and higher bit depth. Screen Content Coding Extensions (SCC) are based on RExt, mainly focusing on improving the coding efficiency for screen content [4]. The development of SCC started in Apr. 2014 and is expected to be finalized in early 2016. Both RExt and SCC are single-layer extensions of HEVC. There are also several extensions of HEVC targeting multiple layers. Scalable HEVC extension (SHVC) focuses on serving a same content with different bandwidth (e.g., different spatial resolution, as known as spatial scalability and different quality, as known as SNR scalability) [5]. Multiview and 3D extensions focus on the encoding of multiple views video content. HEVC Version 2 includes range extensions, scalable extensions and multiview extensions. 3D video coding is enabled in HEVC version 3. SCC will be included in HEVC version 4, which is expected to be finalized in early 2016.

The version 1 of HEVC was finalized in Jan. 2013. Only 4:2:0 chroma sampling format with 8–10 bit per sample was considered in HEVC version 1. To enhance capabilities, HEVC range extensions handle different chroma sampling formats, such as 4:4:4, 4:2:2, and 4:0:0 (monochrome), and higher bit depth encoding. Several new coding tools are added into HEVC range extension, such as cross-component prediction (CCP) [6] and Residual Differential Pulse-Code Modulation (RDPCM) [7], etc. This paper provides an overview of the new added coding tools and comprehensive experimental results comparing with HEVC range extensions with previous video coding standard are also provided.

The rest of this paper is organized as follows. Section 2 introduces HEVC version 1 briefly. Section 3 focuses on the new coding tools in HEVC range extensions. Section 4 provides several experimental results to show the coding efficiency of HEVC range extensions. Section 5 concludes the paper.

## 2 Brief Introduction to HEVC Version 1

Similar to H.264/MPEG-4 AVC, a block-based hybrid framework is applied to HEVC (**Fig. 1**). Intra- or inter-prediction is applied for each block. A 2D transform may be applied to the prediction residue (the other option is transform skip [8], [9], which skips the transform process and the residue is signaled in pixel domain rather than transform domain). The quantized coefficients together with mode information are signaled in the bitstream via Context-Adaptive Binary Arithmetic Cod-



reference picture index(es), and moved from a specific merge candidate. These can be used.

inter prediction direction, reference picture index(es), and motion vector(s)) are inherited from a specific merge candidate. Up to five merge candidates can be used.

#### 4) Transform

4 x 4 to 32 x 32 DCT-like transform can be used in HEVC. For the 4x4 luma TUs in intra-prediction CU, a 4 x 4 DST-like transform is used [13], [14]. A special transform skip mode is also supported for certain type of content (especially screen content) in 4 x 4 TUs [8], [9].

### 3 HEVC Range Extensions

This section gives an overview of HEVC range extensions. New features in HEVC range extensions can be divided into three categories: extension of chroma sampling formats, extension of bit depths, and new coding efficiency enhancement tools.

One of the main purpose of developing HEVC range extensions is to support different chroma sampling formats. Only 4:2:0 is supported in the HEVC version 1 profiles, in which the chroma components have half the resolution of luma in both horizontal and vertical directions. However, a higher chroma fidelity is required in some applications. Besides 4:2:0, the range extensions support 4:4:4 (where the chroma components have the same resolution as luma), 4:2:2 (where the chroma components have the half horizontal resolution as luma, but the same vertical resolution as luma), and 4:0:0 (monochrome, only the video content only has the luma component).

In the 4:4:4 case, the decoding process is quite similar to the 4:2:0 decoding process. The only difference is that the two chroma components have the same spatial resolution as the luma component. One square luma rectangle still corresponds to two square chroma rectangles, the only difference being that all three rectangles are the same size. If 4:4:4 coding is used, the video can be coded in RGB format directly or in YCbCr format. Usually, in the RGB coding, the G component is treated as the luma component and the R and B components are treat-



## An Introduction to High Efficiency Video Coding Range Extensions

Bin Li and Jizheng Xu

ed as the chroma components.

In the 4:2:2 case, the decoding process needs to be changed accordingly, as one square luma block corresponds to two non-square chroma rectangles. For example, a 16 x 16 luma block corresponds to two 8 x 16 chroma blocks (one 8 x 16 Cb block and one 8 x 16 Cr block). To avoid introducing a new non-square transform, the chroma transform needs to be specially handled. In the HEVC range extensions, the non-square chroma block will further split in vertical direction. Thus, two chroma transforms with half the horizontal luma size and half the vertical luma size will be used. In the above example, the 8x16 chroma blocks will be split into 8x8 blocks. So, for each chroma component, two 8 x 8 transforms are applied. The deblocking filter is applied to the newly added transform edge in the 4:2:2 content.

### 3.2 Extension of Bit Depths

The other main purpose of HEVC range extensions is to support higher bit depth encoding. Only up to 10 bit is supported in the HEVC version 1. But some applications, such as those for medical and military purposes, require higher fidelity. Thus, higher bit depth encoding is supported in HEVC range extensions. The main changes to support higher bit depth includes: when extended precision processing is enabled, the dynamic range of coefficients is enlarged and the de-quantization process is adjusted accordingly. When high precision offsets is enabled, the precision of weighted prediction is increased. The SAO offsets can also be scaled up to better support the higher bit depth content.

### 3.3 New Coding Efficiency Enhancement Tools

Several new coding tools are included in the HEVC range extensions to improve the coding efficiency or to provide finer control of encoding parameters. This sub section provides a brief introduction of them.

Cross-Component Prediction (CCP): CCP is used to remove the correlation among color components [6]. CCP is primarily designed for RGB content, but it also provides some bit saving for YCbCr content. CCP is only enabled for 4:4:4 content. When CCP is used, the residue of the first component is used to predict the residue of the other two components via a linear prediction model. The CCP is only used when the three components use the same method to generate the prediction (including inter prediction and intra-prediction if the three components use the same intra-prediction direction, i.e., DM mode for chroma).

Residual Differential Pulse — Code Modulation (RDPCM): Two kinds of RDPCMs are supported in HEVC range extensions [7]. RDPCM modifies the residue in pixel domain, so it is enabled when the residue is signaled in pixel domain. When the transform is bypassed, e.g., in the blocks coding in lossless mode or TUs coded with transform skip, the RDPCM may be used. When horizontal RDPCM is used, the decoded residue is

modified as  $r[x][y] += r[x-1][y]$  and when vertical RDPCM is used, the decoded residue is modified as  $r[x][y] += r[x][y-1]$ , where  $r[x][y]$  is the residue at the (x, y). The residue is modified one by one, and the modification process looks like differential coding. Thus, it is called RDPCM. For intra-coded CUs, implicit RDPCM is used. The horizontal and vertical RDPCM is applied when the horizontal and vertical intra-prediction is used, respectively. For inter-coded CUs, explicit RDPCM is used. The RDPCM direction is signaled in the bitstream when explicit RDPCM is used. Because RDPCM is only enabled for lossless coded blocks and transform skip TUs, it mainly helps to improve the coding efficiency for lossless coding and screen content coding.

Improvements on Transform Skip: HEVC range extensions further improve the transform skip mode to provide better coding efficiency. In HEVC version 1, only 4x4 TUs can use transform skip. In HEVC range extensions, all the TUs, from 4x4 to 32x32, can use transform skip [15]. Rotation is applied to intra 4x4 TUs using transform skip [16]. The coefficients at the right bottom are moved to the upper left, using the equation of  $r[x][y] = \text{coeff}[4-x-1][4-y-1]$ , where  $r[x][y]$  means the rotated coefficient at (x, y) position and  $\text{coeff}[x][y]$  means the unmodified coefficient at (x, y). This technique is also applied to the lossless coded blocks (where transform is bypassed). The context to encode the significant map of transform bypass (including transform skip and lossless coded) TUs is also modified to improve the coding efficiency [16].

Others: The intra reference pixel smoothing filter can be disabled in the HEVC range extensions [17]. Disabling intra reference pixel smoothing filter helps the lossless encoding. Localized control of chroma Quantization Parameter (QP) is supported to provide the ability to adjust the chroma QP in a finer granularity [18]. Several new coding tools are added into the HEVC range extensions, such as persistent rice parameter adaptation [19], CABAC bypass alignment [20], etc.

### 3.4 HEVC Range Extensions Profiles

Several new profiles have been defined for HEVC range extensions. The extended precision processing is enabled in the 16-bit profiles and disabled in the other profiles. CABAC bypass alignment is enabled in High Throughput profile and disabled in all the other profiles.

Monochrome, Monochrome 12 and Monochrome 16 profiles are defined 4:0:0 (monochrome) content with different bit depth range. All the new range extensions coding tools can be enabled in Monochrome 16 profile, but they cannot be used in Monochrome and Monochrome 12 profiles.

Main 12 profile only extends the bit depth range of Main profile to 8–12 bits. 4:2:0 and 4:0:0 contents can be used in Main 12 profile. The new range extensions coding tools in range extensions are not enabled in Main 12 profile.

Main 4:2:2 10 and Main 4:2:2 12 profiles are defined for 4:2:2 content with different bit depth range. 4:2:0 and 4:0:0 con-



tent can also be used in these profiles. All the new range extensions coding tools, except localized control of chroma QP, are disabled in these two profiles.

Main 4:4:4, Main 4:4:4 10 and Main 4:4:4 12 profiles are defined to support 4:4:4 content with different bit depth range. All the chroma sampling formats, including 4:2:0, 4:4:4, 4:0:0, and 4:2:2, can be used in these profiles. All the new range extensions coding tools can be used in these profiles.

Main Intra, Main 10 Intra, Main 12 Intra, Main 4:2:2 10 Intra, Main 4:2:2 12 Intra, Main 4:4:4 Intra, Main 4:4:4 10 Intra, Main 4:4:4 12 Intra and Main 4:4:4 16 Intra profiles are defined for all intra coding.

Main 4:4:4 Still Picture and Main 4:4:4 16 Still Picture profiles are defined for the case there is only one intra picture in the whole bitstream.

High Throughput 4:4:4 16 Intra profile is defined for all intra coding, with CABAC bypass alignment enabled.

## 4 Coding Efficiency of HEVC Range Extensions

To show the coding efficiency of range extensions, this section provides the coding efficiency results of HEVC range extensions with previous video coding standards. The first part of this section compares HEVC range extensions with H.264/MPEG-4 AVC High Predictive profiles. The second part of this section compares HEVC range extensions with HEVC version 1. The latest available reference software is used in the test. HM-16.7 [21] is used to generate HEVC version 1 and range extensions bitstreams and JM-19.0 [22] is used to generate H.264/MPEG-4 AVC bitstreams. Both HM-16.7 and JM-19.0 are configured with similar settings.

Three coding structures are used in the tests. One of these is Random Access (RA) coding structure, in which intra refresh is relatively frequent and the delay is not a critical issue. In the test, random access points are inserted into the bitstreams about once a second. A Hierarchical - B coding structure with group of pictures (GOP) size of 8 is used in the RA coding structure. Temporal scalability with four different layers is supported in the HEVC RA coding structure, while it is not supported in the H.264/MPEG-4 AVC RA coding structure. The supporting of temporal scalability with four different temporal layers in HEVC RA coding structure brings about 0.3% performance drop on average [23]. Besides, the low delay (LD) B coding structure is used for real-time communications, in which the coding delay is critical and the random access support is less important. IBBB (without picture reordering) coding structure with hierarchical quantization parameter (QP) is used in LD coding. The third one is all-intra (AI) coding structure, in

which no temporal prediction is applied and all the pictures use intra-picture prediction only.

Only objective PSNR-based test results are provided in this section. The coding efficiency is measured in terms of Bjøntegaard-delta bit rate (BD-rate) [24], which measures the bit rate difference at the same quality. A negative number means bit rate reduction (performance gain) and a positive number means bit rate increase (performance loss).

### 4.1 Comparison of HEVC Range Extensions with H.264/MPEG-4 AVC High 4:4:4 Predictive Profile

To show the coding efficiency of HEVC RExt, we compare it with H.264/MPEG-4 AVC High 4:4:4 Predictive profile. Two sets of coding results are provided in this paper. The first test set uses the sequences specified in HEVC RExt Common Test Condition (CTC) [25]. The sequences in the first test set are 8–12 bit per sample, in YUV 4:2:2, YUV 4:4:4 and RGB 4:4:4 format. The second test set uses the Netflix sequence [26], which is in YUV 4:4:4 10-bit format, with a spatial resolution of 4096 x 2160 and the temporal resolution 60 Hz to reflect the 4K video application. We choose 10 clips (120 pictures in each clip) from the Netflix sequence to conduct the test. The start time in the original sequence of the 10 clips is provided

▼ Table 1. Clips of Netflix sequence used in the test

Clip name	Start time (s)	Clip name	Start time(s)
NarrotorWorking	6.84	CityDayView	27.76
Vegetable	39.44	FlowerMarket	46.30
Vegetable2	55.34	FoodMarket	70.53
PeopleWalking	81.89	AztecRitualDance	97.70
CouplesDancing	115.27	Motorcycles	133.00

▼ Table 2. Coding performance of HEVC range extensions over H.264/MPEG-4 AVC (RExt CTC sequences)

	All Intra Main-tier			All Intra High-tier			All Intra Super-High-tier		
	Y/G	U/B	V/R	Y/G	U/B	V/R	Y/G	U/B	V/R
RGB 4:4:4	−34.7%	−27.2%	−29.4%	−28.0%	−24.2%	−25.6%	−23.2%	−20.2%	−21.3%
YCbCr 4:4:4	−26.4%	−26.0%	−30.9%	−25.0%	−26.9%	−33.6%	−22.5%	−27.1%	−33.9%
YCbCr 4:2:2	−21.4%	−13.5%	−14.3%	−18.1%	−13.9%	−17.7%	−14.3%	−12.1%	−15.3%

	Random Access Main-tier			Random Access High-tier		
	Y/G	U/B	V/R	Y/G	U/B	V/R
RGB 4:4:4	−40.1%	−35.5%	−36.3%	−32.3%	−30.3%	−31.2%
YCbCr 4:4:4	−40.0%	−51.2%	−50.1%	−38.8%	−47.9%	−56.8%
YCbCr 4:2:2	−31.9%	−21.7%	−20.4%	−28.3%	−28.1%	−30.4%

	Low Delay B Main-tier			Low Delay B High-tier		
	Y/G	U/B	V/R	Y/G	U/B	V/R
RGB 4:4:4	−39.8%	−35.1%	−37.2%	−30.9%	−30.6%	−31.3%
YCbCr 4:4:4	−45.2%	−56.1%	−62.2%	−42.0%	−49.3%	−60.4%
YCbCr 4:2:2	−37.7%	−28.4%	−28.4%	−32.8%	−30.7%	−35.3%

## An Introduction to High Efficiency Video Coding Range Extensions

Bin Li and Jizheng Xu

▼ **Table 3. Coding performance of HEVC range extensions over H.264/MPEG-4 AVC (Netflix sequence clips)**

	All Intra Main-tier			All Intra High-tier			All Intra Super-High-tier		
	Y/G	U/B	V/R	Y/G	U/B	V/R	Y/G	U/B	V/R
AztecRitualDance	-25.5%	-25.5%	-34.1%	-25.1%	-34.1%	-45.5%	-24.6%	-44.4%	-58.4%
CityDayView	-27.1%	-32.5%	-42.4%	-26.5%	-40.9%	-49.9%	-23.4%	-44.5%	-60.1%
CouplesDancing	-46.7%	-60.8%	-59.6%	-47.7%	-68.1%	-71.4%	-36.5%	-76.3%	-81.2%
FlowerMarket	-27.9%	-31.3%	-34.6%	-22.2%	-38.5%	-42.9%	-22.5%	-43.7%	-49.3%
FoodMarket	-31.4%	-27.7%	-42.8%	-29.4%	-32.6%	-52.3%	-25.6%	-31.1%	-58.8%
Motorcycles	-52.8%	-52.4%	-58.5%	-51.8%	-63.5%	-71.5%	-37.1%	-74.7%	-85.2%
NarrotorWorking	-35.8%	-39.6%	-37.8%	-38.2%	-44.9%	-43.5%	-39.6%	-47.6%	-45.2%
PeopleWalking	-52.8%	-60.2%	-62.2%	-55.2%	-70.3%	-76.8%	-38.4%	-79.1%	-88.5%
Vegetable	-24.0%	-22.7%	-30.0%	-20.0%	-26.7%	-34.8%	-18.4%	-23.9%	-28.2%
Vegetable2	-37.2%	-37.4%	-42.8%	-34.8%	-47.2%	-53.6%	-28.0%	-59.0%	-66.0%
Average	-36.1%	-39.0%	-44.5%	-35.1%	-46.7%	-54.2%	-29.4%	-52.4%	-62.1%

	Random Access Main-tier			Random Access High-tier		
	Y/G	U/B	V/R	Y/G	U/B	V/R
AztecRitualDance	-44.6%	-53.8%	-59.3%	-45.3%	-63.9%	-70.9%
CityDayView	-58.8%	-73.8%	-73.2%	-63.3%	-84.6%	-89.9%
CouplesDancing	-75.0%	-87.0%	-83.1%	-76.8%	-94.6%	-92.0%
FlowerMarket	-57.8%	-61.5%	-71.9%	-33.6%	-50.8%	-59.1%
FoodMarket	-59.4%	-58.2%	-76.1%	-52.4%	-54.7%	-85.7%
Motorcycles	-81.2%	-84.8%	-85.6%	-86.4%	-93.5%	-95.0%
NarrotorWorking	-62.4%	-76.9%	-70.8%	-62.7%	-86.8%	-81.8%
PeopleWalking	-74.3%	-83.7%	-81.5%	-77.5%	-92.1%	-91.6%
Vegetable	-52.5%	-59.9%	-70.8%	-30.2%	-46.0%	-60.3%
Vegetable2	-66.5%	-76.5%	-76.6%	-69.1%	-85.9%	-88.4%
Average	-63.2%	-71.6%	-74.9%	-59.7%	-75.3%	-81.5%

	Low Delay B Main-tier			Low Delay B High-tier		
	Y/G	U/B	V/R	Y/G	U/B	V/R
AztecRitualDance	-51.5%	-65.2%	-68.8%	-50.4%	-67.7%	-73.9%
CityDayView	-75.9%	-87.8%	-90.5%	-73.9%	-89.7%	-96.7%
CouplesDancing	-77.6%	-89.0%	-85.0%	-77.6%	-92.4%	-90.6%
FlowerMarket	-56.2%	-62.6%	-69.8%	-35.3%	-54.3%	-60.0%
FoodMarket	-61.9%	-65.0%	-85.2%	-51.5%	-58.3%	-87.7%
Motorcycles	-81.4%	-86.3%	-87.3%	-83.7%	-91.1%	-93.4%
NarrotorWorking	-69.5%	-84.8%	-79.9%	-67.2%	-86.5%	-82.9%
PeopleWalking	-76.4%	-84.7%	-83.1%	-78.7%	-89.4%	-89.9%
Vegetable	-56.1%	-64.6%	-78.6%	-34.7%	-50.1%	-64.3%
Vegetable2	-75.3%	-86.3%	-87.5%	-73.1%	-88.1%	-91.6%
Average	-51.5%	-65.2%	-68.8%	-50.4%	-67.7%	-73.9%

in **Table 1**. The coding results using HEVC CTC sequences are provided in **Table 2** and the coding results using Netflix sequences are provided in **Table 3**. The QP range is 22–37 for Main-tier, 17–32 for High-tier, and 12–27 for Super-High-tier.

From Table 2, we can know that for RGB 4:4:4 sequences,

compared with H.264/MPEG - 4 AVC, HEVC saves 23.2%–34.7% bit-rate for All Intra coding. 32.3%–40.1% bits saving is achieved for Random Access coding and 30.9%–39.8% bits saving is achieved for Low Delay B coding, at different bit rate ranges. Table 2 also shows that the bit saving is higher at Main-tier, which indicates that improving the coding efficiency at high quality end is more challenging.

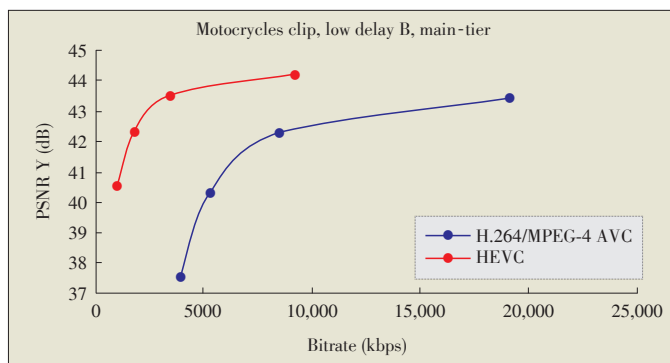
It can be seen from Table 3 that, when compared with H.264/MPEG - 4 AVC, HEVC saves about 29.4%–36.1% bits for All Intra coding. 59.7%–63.2% bits saving is achieved for Random Access coding and 50.4%–51.5% bits saving is achieved for Low Delay B coding. Table 3 also shows that the bits saving of HEVC of H.264/MPEG - 4 AVC is much larger for 4K sequences. An example R-D curve of Motorcycles clip under Low Delay B coding structure at Main - tier is shown in **Fig. 2**.

### 4.2 Comparison of HEVC Range Extensions with HEVC Version 1

We also provide the coding efficiency of HEVC RExt over HEVC version 1. We use HEVC version 1 test sequences and HEVC version 1 CTC [27] to perform the test. The overall coding performance of HEVC RExt over HEVC version 1 is provided in **Table 4**. The QP range used in the test is 22–37. Both HEVC version 1 encoding and RExt encoding are configured using 4:2:0 8-/10-bit encoding. The only difference is that new coding tools are enabled in the RExt en-

coding configurations.

Table 4 shows that for 4:2:0 content, HEVC range extensions do not provide much performance improvements except for Class F sequences. The main reason for this phenomenon is that Class F sequences are screen content and HEVC range ex-



▲ Figure 2. R-D curve of motorcycles clip under low delay B coding structure at main-tier.

▼ Table 4. Coding performance of HEVC RExt over HEVC version 1

	All Intra Main			All Intra Main10		
	Y	U	V	Y	U	V
Class A	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Class B	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Class C	−0.3%	−0.2%	−0.3%	−0.3%	−0.2%	−0.3%
Class D	−0.5%	−0.5%	−0.6%	−0.5%	−0.4%	−0.5%
Class E	−0.2%	−0.1%	−0.1%	−0.2%	0.0%	−0.1%
Class F	−4.9%	−5.2%	−5.4%	−5.0%	−5.2%	−5.7%
Overall	−1.0%	−1.0%	−1.1%	−0.2%	−0.1%	−0.2%

Random Access Main				Random Access Main10		
Class A	0.0%	−0.1%	0.3%	0.0%	0.3%	−0.4%
Class B	0.0%	0.1%	0.1%	0.0%	−0.1%	−0.1%
Class C	−0.2%	−0.4%	−0.4%	−0.2%	−0.5%	−0.6%
Class D	−0.3%	−0.5%	−0.7%	−0.3%	−0.5%	−0.5%
Class E						
Class F	−3.7%	−4.2%	−4.3%	−3.8%	−3.9%	−4.6%
Overall	−0.1%	−0.2%	−0.2%	−0.1%	−0.2%	−0.4%

Low Delay B Main			Low Delay B Main10			
Class A						
Class B	0.0%	0.1%	0.0%	0.0%	0.2%	0.4%
Class C	−0.1%	0.1%	0.1%	−0.1%	0.2%	0.1%
Class D	−0.1%	0.2%	0.1%	0.0%	0.1%	0.6%
Class E	−0.1%	−1.1%	2.1%	−0.1%	−0.3%	1.0%
Class F	−2.4%	−2.1%	−2.3%	−2.5%	−2.4%	−4.2%
Overall	−0.5%	−0.5%	−0.1%	−0.5%	−0.4%	−0.4%

Low Delay P Main				Low Delay P Main10		
Class A						
Class B	0.0%	0.0%	0.0%	0.0%	0.0%	0.4%
Class C	−0.1%	0.1%	0.2%	0.0%	0.3%	0.0%
Class D	0.0%	0.2%	0.6%	0.0%	0.3%	0.8%
Class E	−0.2%	−1.3%	1.3%	−0.1%	−0.4%	0.8%
Class F	−2.5%	−1.8%	−2.2%	−2.4%	−2.4%	−3.5%
Overall	−0.5%	−0.5%	−0.1%	−0.5%	−0.4%	−0.3%

tensions improves quite a lot for screen content. For nature content, HEVC range extensions provide almost the same coding efficiency as HEVC version 1.

## 5 Conclusion

This paper provides an overview of HEVC range extensions. HEVC range extensions provide the ability to handle higher bit depths and higher fidelity chroma sampling formats for video. Several new coding tools are also added in the HEVC range extensions. The experimental results show that for 4K sequences, compared with H.264/MPEG-4 AVC High Predictive profile, HEVC range extensions save about 36.1% bit-rate for All intra-coding, 63.2% bit-rate for Random Access coding and 51.5% bit-rate for Low Delay B coding, at Main-tier quality range.

## References

- [1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012. doi: 10.1109/TCSVT.2012.2221191.
- [2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, Jul. 2003. doi: 10.1109/TCSVT.2003.815165.
- [3] J. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012. doi: 10.1109/TCSVT.2012.2221192.
- [4] ITU-T Q6/16 Visual Coding and ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, "Joint Call for Proposals for Coding of Screen Content," 49th VCEG Meeting, San José, USA, document VCEG-AW90, 2014.
- [5] G. J. Sullivan, J. M. Boyce, C. Ying, 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.
- [6] W. Pu, W.-K. Kim, J. Chen, et al., "Non-RCE1: inter color component residual prediction," 14th JCT-VC meeting, Vienna, Austria, document JCTVC-N0266, 2013.
- [7] R. Joshi, J. Sole, and M. Karczewicz, "RCE2 subtest C.2: extension of residual DPCM to lossy coding," 14th JCT-VC meeting, Vienna, Austria, document JCTVC-N0052, 2013.
- [8] C. Lan, J. Xu, G. J. Sullivan, and F. Wu, "Intra transform skipping," 9th JCT-VC meeting, Geneva, Switzerland, document JCTVC-I0408, 2012.
- [9] X. Peng, C. Lan, J. Xu, and G. J. Sullivan, "Inter transform skipping," 10th JCT-VC meeting, Stockholm, Sweden, document JCTVC-J0237, 2012.
- [10] C.-M. Fu, E. Alshina, A. Alshin, et al., "Sample adaptive offset in the HEVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1755–1764, Dec. 2012. doi: 10.1109/TCSVT.2012.2221529.
- [11] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J. Park, "Block partitioning structure in the HEVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1697–1706, Dec. 2012. doi: 10.1109/TCSVT.2012.2223011.
- [12] K. Ugur, A. Alshin, E. Alshina, et al., "Motion compensated prediction and interpolation filter design in H.265/HEVC," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 946–956, Dec. 2013. doi: 10.1109/JSTSP.2013.2272771.
- [13] M. Budagavi, A. Fuldseth, G. Bjontegaard, V. Sze, and M. adafale, "Core transform design in the high efficiency video coding (HEVC) standard," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1029–1041, Dec. 2013. doi: 10.1109/JSTSP.2013.2270429.

## An Introduction to High Efficiency Video Coding Range Extensions

Bin Li and Jizheng Xu

- [14] J. Sole, R. Joshi, N. Nguyen, *et al.*, "Transform coefficient coding in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1765–1777, Dec. 2012. doi: 10.1109/TCSVT.2012.2223055.
- [15] X. Peng, J. Xu, B. Li, *et al.*, "Non-RCE2: transform skip on large TUs," 14th JCTVC meeting, Vienna, Austria, document JCTVC-N0288, 2013.
- [16] J. Sole, R. Joshi, and M. Karczewicz, "RCE2 Test B.1: residue rotation and significance map context," 14th JCT-VC meeting, Vienna, Austria, document JCTVC-N0044, 2013.
- [17] J. Zhu, and K. Kazui, "Non-RCE2: skip of neighbouring samples filtering in intra-prediction for lossless coding," 14th JCT-VC meeting, Vienna, Austria, document JCTVC-N0080, 2013.
- [18] D. Flynn, N. Nguyen, D. He, *et al.*, "RExt: CU-adaptive chroma QP offsets," 15th JCT-VC meeting, Geneva, Switzerland, document JCTVC-00044, 2013.
- [19] M. Karczewicz, L. Guo, J. Sole, *et al.*, "RCE2: results of Test 1 on rice parameter initialization," 16th Meeting, San Jose, USA, document JCTVC-P0199, 2014.
- [20] K. Sharman, N. Saunders, and J. Gamei, "RCE1: results of tests B1, B2 and B3a," 16th Meeting, San Jose, USA, document JCTVC-P0060, 2014.
- [21] HEVC Model (HM) Reference Software Versions HM-16.7 [Online]. Available: [https://hevc.hhi.fraunhofer.de/svn/svn\\_HEVCSoftware/tags/](https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/)
- [22] H.264/MPEG-4 AVC Joint Model (JM) Reference Software Version JM 18.6 [Online]. Available: <http://iphome.hhi.de/suehring/tml/download/>
- [23] B. Li and J. Xu, "On referencing structure supporting temporal scalability," 18th JCT-VC meeting, Sapporo, Japan, document JCTVC-R0103, 2014.
- [24] G. Bjøntegaard, "Improvements of the BD-PSNR model," 35th VCEG meeting, Berlin, Germany, document VCEG-A111, 2008.
- [25] C. Rosewarne, K. Sharman, and D. Flynn, "Common test conditions and software reference configurations for HEVC range extensions," 16th JCT-VC meeting, San Jose, USA, document JCTVC-P1006, 2014.
- [26] A. Norkin, I. Katsavounidis, A. Aaron, and J. De Cock, "Netflix test sequences for next generation video coding," 56th VCEG meeting, Warsaw, Poland, document VCEG-AZ09, 2015.
- [27] F. Bossen, "Common test conditions and software reference configurations," 12th JCT-VC meeting, Geneva, Switzerland, document JCTVC-L1100, 2013.

Manuscript received: 2015–11–15

## Biographies

**Bin Li** (libin@microsoft.com) received the B.S. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 2008 and 2013, respectively. He joined Microsoft Research Asia (MSRA), Beijing, China, as an Associate Researcher in 2013. He has authored or co-authored over 20 papers. He holds over 10 granted or pending U.S. patents in the area of image and video coding. He has more than 30 technical proposals that have been adopted by Joint Collaborative Team on Video Coding. His current research interests include video coding, processing, and communication. Dr. Li received the best paper award for the International Conference on Mobile and Ubiquitous Multimedia from Association for Computing Machinery in 2011. He received the Top 10% Paper Award of 2014 IEEE International Conference on Image Processing. He has been an active contributor to ISO/MPEG and ITU-T video coding standards. (JCT-VC). He is currently the Co-Chair of the Ad Hoc Group of Screen Content Coding extensions software development.

**Jizheng Xu** (jzxu@microsoft.com) (M'07-SM'10) received the B.S. and M. S. degrees in computer science from the University of Science and Technology of China (USTC), and the Ph.D. degree in electrical engineering from Shanghai Jiaotong University, China. He joined Microsoft Research Asia (MSRA) in 2003 and currently he is a Lead Researcher. He has authored and co-authored over 100 conference and journal refereed papers. He has over 30 U.S. patents granted or pending in image and video coding. His research interests include image and video representation, media compression, and communication. He has been an active contributor to ISO/MPEG and ITU-T video coding standards. He has over 40 technical proposals adopted by H.264/AVC, H.264/AVC scalable extension, High Efficiency Video Coding, HEVC range extension and HEVC screen content coding standards. He chaired and co-chaired the ad-hoc group of exploration on wavelet video coding in MPEG, and various technical ad-hoc groups in JCT-VC, e.g., on screen content coding, on parsing robustness, on lossless coding. He co-organized and co-chaired special sessions on scalable video coding, directional transform, high quality video coding at various conferences. He also served as special session co-chair of IEEE International Conference on Multimedia and Expo 2014.

## Call for Papers

### ZTE Communications Special Issue on Multi-Gigabit Millimeter-Wave Wireless Communications

The exponential growth of wireless devices in recent years has motivated the exploration of the millimeter-wave frequency spectrum for multi-gigabit wireless communications. Recent advances in antenna technology, RF CMOS process, and high-speed baseband signal processing algorithms make millimeter-wave wireless communication feasible. The multi-gigabit-per-second data rate of millimeter-wave wireless communication systems will lead to applications in many important scenarios, such as WPAN, WLAN, back-haul for cellular system. The frequency bands include 28 GHz, 38 GHz, 45GHz, 60GHz, E-BAND, and even beyond 100 GHz. The upcoming special issue of *ZTE Communications* will present some major achievements of the research and development in multi-gigabit millimeter-wave wireless communications. The expected publication date will be in December 2016. It includes (but not limited to) the following topics:

- Channel characterization and channel models
- Antenna technologies
- Millimeter-wave-front-end architectures and circuits

- Baseband processing algorithms and architectures
- System aspects and applications

#### Paper Submission

Please directly send to [eypzhang@ntu.edu.sg](mailto:eypzhang@ntu.edu.sg) and use the email subject "ZTE-MGMMW-Paper-Submission".

#### Tentative Schedule

Paper submission deadline: June 15, 2016

Editorial decision: August 31, 2016

Final manuscript: September 15, 2016

#### Guest Editors

Prof. Yueping Zhang, Nanyang Technological University, Singapore ([eypzhang@ntu.edu.sg](mailto:eypzhang@ntu.edu.sg))

Prof. Ke Guan, Beijing Jiao Tong University, China ([kguan@bjtu.edu.cn](mailto:kguan@bjtu.edu.cn))

Prof. Junjun Wang, Beihang University, China ([wangjunjun@buaa.edu.cn](mailto:wangjunjun@buaa.edu.cn))

# Multi-Layer Extension of the High Efficiency Video Coding (HEVC) Standard

Ming Li and Ping Wu

(ZTE Corporation, Shenzhen 518057, China)

## Abstract

Multi-layer extension is based on single-layer design of High Efficiency Video Coding (HEVC) standard and employed as the common structure for scalability and multi-view video coding extensions of HEVC. In this paper, an overview of multi-layer extension is presented. The concepts and advantages of multi-layer extension are briefly described. High level syntax (HLS) for multi-layer extension and several new designs are also detailed.

## Keywords

HEVC; multi-layer extension

## 1 Introduction

High Efficiency Video Coding (HEVC) standard is the newest video coding standard of the ITU-T Q6/16 Video Coding Experts Group (VCEG) and the ISO/IEC JTC 1 SC 29/WG 11 Moving Picture Experts Group (MPEG). The first version of HEVC standard was released in 2013 [1] and referred to as “HEVC Version 1” standard. It is the next generation video coding standard after H.264/AVC, and achieves a dramatic improvement of coding efficiency relative to existing H.264/AVC. Testing results demonstrate that HEVC brings the same subjective quality by consuming an average of 50% fewer coding bits than that of H.264/AVC [2], [3]. HEVC standard is believed to be adopted in most of the potential applications employing video coding including broadcast, storage, streaming, surveillance, video telephony and etc.

To address the requirements of a wider range of applications, key extensions of the HEVC Version 1 standard have been introduced by the Joint Collaboration Team on Video Coding (JCT-VC) and Joint Collaboration Team on 3D Video Coding Extension Development (JCT-VC) of VCEG and MPEG [4]. Range extensions (RExt), multiview extension (MV-HEVC) and scalable extension (SHVC) were introduced and included in the second version of HEVC standard [5]. 3D high-efficiency video coding extension (3D-HEVC) was finalized as the latest extension in the third version of HEVC standard [6] to enable high-coding of the representative 3D video signal of “multiview video + multiview depth”. Currently, new extensions for Screen Content Coding (SCC) [7] and high dynamic range and

wide color gamut (HDR & WCG) [8] are being developed in VCEG and MPEG and are scheduled for release in the coming one or two years.

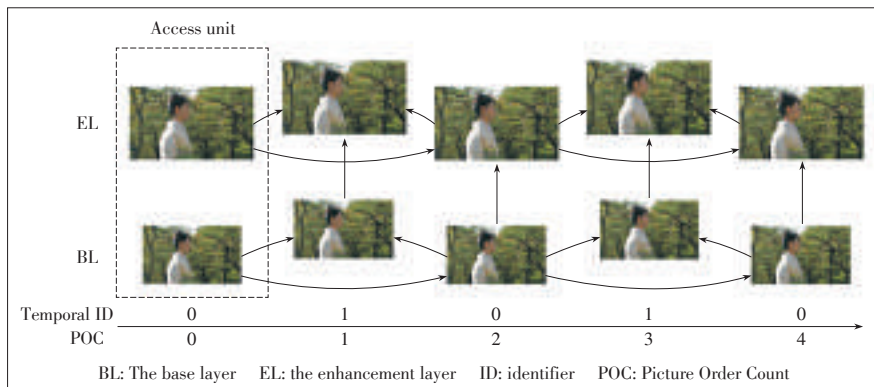
In the second version of the HEVC standard, the concept of layer refers to a scalable layer (e.g. a spatial scalable layer) in SHVC or a view in MV-HEVC. **Fig. 1** shows an example SHVC bitstream of spatial scalability with two layers. The base layer (BL) is of lower resolution, and the enhancement layer (EL) higher resolution. In both BL and EL, temporal scalability, which is already supported in HEVC Version 1, is ensured by using hierarchical B-pictures, and the pictures with temporal identifier (ID) equal to 0 and 1 form sub-layers of BL and EL, respectively. A similar layer concept is also applied to the MV-HEVC bitstream in **Fig. 2**, where a layer corresponds to a view and one base view and two dependent views are referred to as BL (central view), EL1(right view) and EL2 (left view), respectively. In both SHVC and MV-HEVC, BL provides backward compatibility to single layer HEVC codec.

For both SHVC and MV-HEVC, the inter-layer prediction is the key to superior coding efficiency compared with simulcast. MV-HEVC is based on HEVC Version 1 standard and follows the same principle of multiview video coding (MVC) extension of H.264/AVC [9], which does not introduce changes to block-level algorithms. In MV-HEVC, inter-layer prediction is carried out by high-level operations to put the reconstructed pictures from reference views to the reference lists of the current picture. Block level tools are adopted in 3D-HEVC for further improving coding efficiency. In the development of SHVC in JCT-VC, in-depth study and testing has been conducted to evaluate the overall performance of two inter-layer prediction

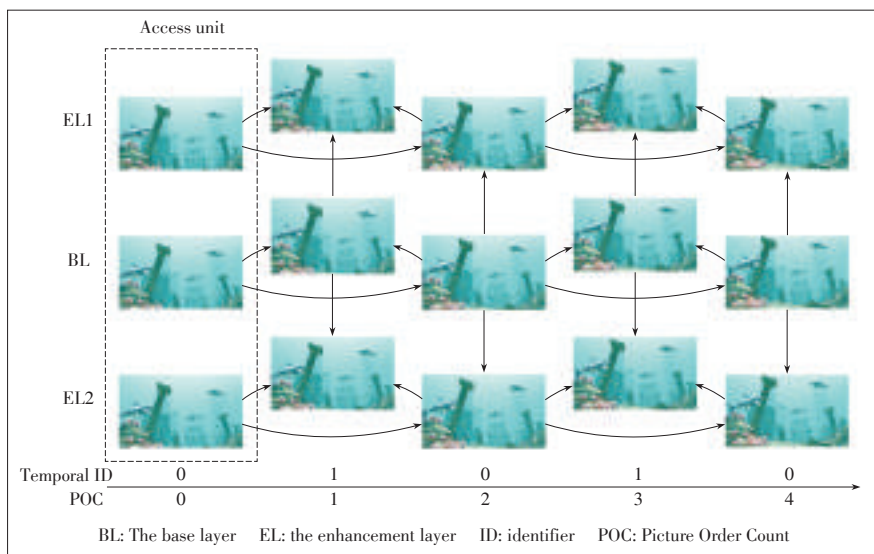


## Multi-Layer Extension of the High Efficiency Video Coding (HEVC) Standard

Ming Li and Ping Wu



▲ Figure 1. An example of SHVC bitstream.



▲ Figure 2. An example of MV-HEVC bitstream.

schemes of high-level extension using reference index to signal inter-layer reference and block level extension with dedicated tools for coding EL [10]. Considering the trade-offs among design advantage, coding efficiency and complexity, JCT-VC selects high-level extension approach for SHVC. Therefore, a multi-layer extension of HEVC is established and employed as the common structure for both MV-HEVC and SHVC, as well as other future extensions using the layered structure.

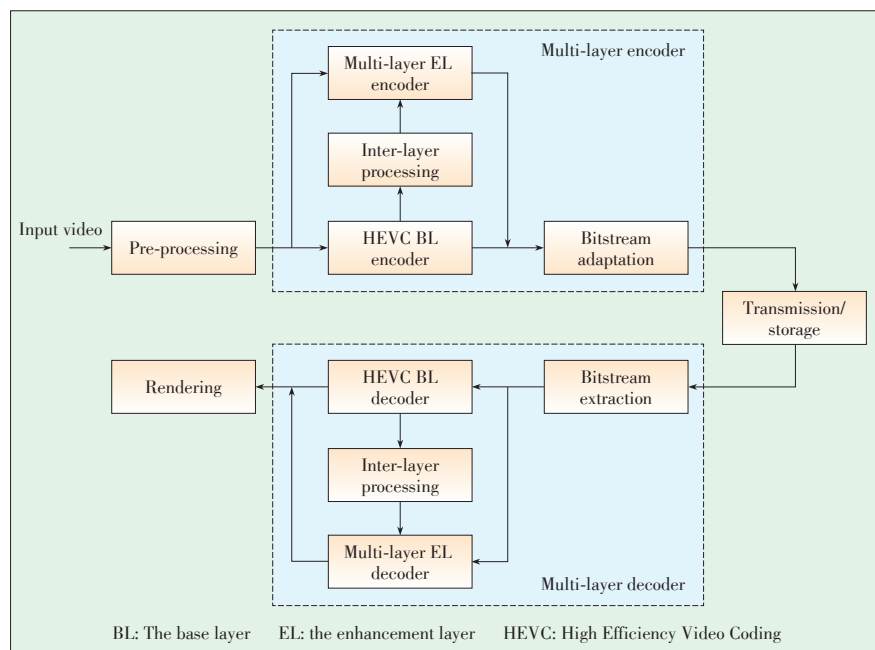
## 2 Multi-Layer Extension

Multi-layer extension represents scalable and multiview structures by layers and offers flexibility to combinations of different types of layered structures. Compared to Scalable Video Coding (SVC) and MVC extensions of H.264/AVC standard, multi-layer extension provides an identical framework for SHVC and MV-HEVC extensions, as well as future extensions employing multi-layer structure, of HEVC standard. From the perspective of decoding operations, the main feature is that the decoding operations at block level are kept the same as those

specified for single layer profiles of HEVC standard. For example, an SHVC decoder conforming to Scalable Main profile is implemented using the same block level decoding algorithms as the ones specified in HEVC Main Profile. In multi-layer extension, inter-layer prediction is enabled by putting the reconstructed pictures of lower layers into the reference lists of the pictures in higher layers within the same access unit.

An example of an end-to-end system employing multi-layer bitstream with a BL and one EL is shown in Fig. 3. At the source side, the pre-processing module is used to get the BL video and EL video for BL encoder and EL encoder, respectively. For example, when the multi-layer encoder is an SHVC encoder of spatial scalability with two layers, the pre-processing module generates a lower resolution video for BL encoder by down-sampling the input video. When the multi-layer encoder is an MV-HEVC encoder with its input of stereo video of two views, the pre-processing module will choose one view as BL and the other as EL according to the configurations. The bitstream adaptation module forms the multi-layer bitstream by combining the coding bits of BL and EL following the specifications of multi-layer extension. At the destination side, the bitstream extraction module in multi-layer decoder is to separate BL and EL stream from the received multi-layer bitstream, e.g. by running the bitstream extraction process. The rendering module at destination is to show the decoded video according to the requests from users. In the above mentioned examples, the rendering module displays the desired video from an SHVC decoder of spatial scalability, or constructs a stereo pair from the decoded videos from an MV-HEVC decoder for 3D viewing.

At both sides of source and destination, when inter-layer prediction is used, the inter-layer processing module accesses the decoded picture buffer (DPB) of the BL encoder (or BL decoder) to get the corresponding reconstructed BL picture to generate the inter-layer reference picture for encoding (or decoding) the EL picture in the same access unit. When one or more parameters of resolution, bit depth and colour gamut of the BL reconstructed picture are different from the parameters of EL picture, the inter-layer processing module performs necessary operations on the BL reconstructed picture that may include conversions of texture, color and motion field. The output picture of the inter-layer processing module is then put into the inter-layer reference picture set and marked as "used for long-term reference" in encoding (or decoding) the EL picture. In the pro-



▲ Figure 3. End-to-end structure of system employing multi-layer bitstream.

cess of constructing the reference picture lists for the EL picture, the inter-layer reference picture is added to the reference picture list, and assigned with a reference index along with the temporal reference pictures of the EL picture. In HEVC multi-layer extension, the parameters for inter-layer processing and inter-layer prediction are signalled in parameter set and slice segment header. Inter-layer prediction is signalled by setting the values of the syntax elements of reference index in prediction unit (PU) equal to the corresponding reference index of the inter-layer reference picture, and carried out without changing any operations below slice level specified in HEVC Version 1 standard. This is referred to as “high level syntax (HLS) extension scheme”.

The multi-layer codec is of a multi-loop coding structure. The major advantage, especially compared to SVC employing a single-loop design, is the HLS extension scheme that reuses the block level algorithms already designed for HEVC Version 1 codec. The additional operations newly introduced to the EL codec is to interpret the dependency among layers for inter-layer prediction and the inter-layer processing to generate the inter-layer reference picture to be involved in reference lists for EL pictures. Accordingly, the EL codec needs to access the DPB of BL codec for BL reconstructed picture and maybe also the associated motion information of BL picture to derive motion predictor for EL PUs. As the interface for BL motion information already exists for motion prediction at BL, EL can reuse such interface to get BL motion information when an EL PU referencing to the inter-layer reference picture for motion prediction. In this way, a multi-layer codec can be conveniently designed and implemented, for example, by taking the already existing HEVC Version 1 codec as BL codec and inte-

grating an inter-layer processing module as well as high-level interpretation for multi-layer structure signalled in parameter sets and slice segment header in HEVC Version 1 codec to form EL codec. In comparison with SVC codec, the HLS extension scheme avoids a completely new design of EL by reusing most parts of the HEVC Version 1 design, and also saves a large amount of extra interfaces to be implemented on already available single layer design to meet EL’s accessing. Therefore, the HLS extension scheme greatly brings down the workload for SHVC and MV-HEVC codec design and implementation, which is believed to push wide adoption of layered coding extensions of HEVC to applications.

### 3 HLS for Multi-Layer Extension

To describe the common layered structure of SHVC and MV-HEVC, HLS specified in HEVC Version 1 standard is further extended, including network abstraction layer (NAL) unit header, parameter sets, slice segment header and supplement enhancement information (SEI). New designs are being introduced to make the multi-layer extension more flexible for applications and future extensions using layered structure.

#### 3.1 NAL Unit Header and Parameter Sets

Multi-layer extension shares the same NAL unit header as that specified in HEVC Version 1 standard. In NAL unit header, a syntax element namely `nuh_layer_id` is coded in 6 bits to signal the layer to which a video coding layer (VCL) NAL unit or non-VCL NAL unit belongs to. In HEVC Version 1 standard, the value of `nuh_layer_id` in the conformed bitstream shall be 0 and the conformed decoder ignores all NAL units with `nuh_layer_id` not equal to 0. In the multi-layer extension, the value of `nuh_layer_id` is always 0 in the BL NAL units, which are backward-compatible with HEVC Version 1 codec. With `nuh_layer_id` distinguishing the NAL units of different layers, the NAL unit types defined in HEVC Version 1 standard are re-used to indicate the type of raw byte sequence payload (Rbsp) data structure contained in the EL NAL units and signalled by the existing syntax element `nuh_unit_type` in NAL header. Therefore, no new NAL types are introduced by the multi-layer extension.

Video parameter set (VPS) is adopted in the development of HEVC Version 1 standard. In multi-layer extension, VPS is further extended to signal the common information for the layers. VPS could be used in session negotiation to provide the characteristics of the multi-layer bitstream and decoding capability. Layers indicated by VPS can be a spatial/quality scalable layer, a view, and an auxiliary picture layer. VPS de-

## Multi-Layer Extension of the High Efficiency Video Coding (HEVC) Standard

Ming Li and Ping Wu

scribes the number of layers and dependency relationship among the layers. The dependency relationship indicates the reference layers for inter-layer prediction when decoding the current layer. In multi-layer extension, a layer can only reference lower layers. VPS signals the representation format for each layer. VPS provides the information for bitstream conformance and operation points, including profile, tier, level, layer sets, hypothetical reference decoder (HRD) parameters, and etc. Video usability information (VUI) for multi-layer bitstream is also signalled in VPS.

Extension of sequence parameter set (SPS) for multi-layer extension introduces only one syntax element `inter_view_mv_vert_constraint_flag`, which is to signal whether the vertical components of motion vectors used for inter-layer prediction are constrained. Extension of picture parameter set (PPS) for multi-layer extension includes parameters for picture processing to derive inter-layer reference pictures, including reference picture scaling offsets, reference region, reference phase offsets and colour mapping.

### 3.2 Layer-Wise Decoding and Picture Order Count (POC) Resetting

In SVC, the decoding process can only correctly start from an access unit with all pictures coded as instantaneous decoding refresh (IDR) pictures. At the encoder side, coding an access unit with all IDR pictures always leads to an instantaneous bit-rate increment. By comparison, multi-layer extension releases the constraint that the intra random access point (IRAP) pictures are aligned within one access unit. A device can start decoding a multi-layer bitstream from an access unit with the BL picture being a random access picture, and make an access to EL layers later, for example, when a random access picture is in the EL layer.

Fig. 4 shows a multi-layer bitstream structure supporting layer-wise decoding. The access unit AU(t4) is an access unit with BL picture coded as a broken link access (BLA) picture. The multi-layer decoder can make an access to the BL of this multi-layer bitstream from AU(t4) first, and then access the EL from AU(t8) in which the EL picture is an IDR picture. Note that in the example shown in Fig. 4, the EL picture in AU(t4) is not decodable, because its temporal reference picture in EL (i.e. the EL picture in AU(t0)) is not available when accessing from AU(t4).

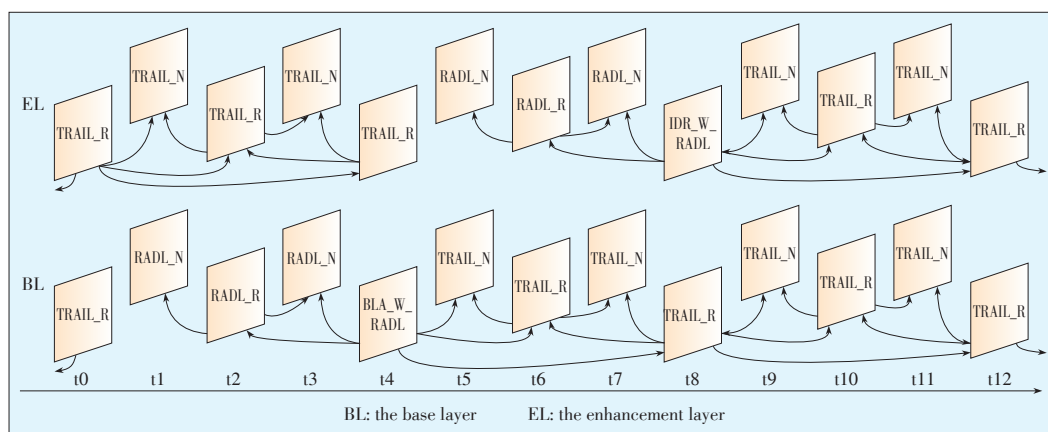
As with the HEVC Version 1 standard, POC is used in multi-layer extension to represent the relative output order of pictures within each

layer, and generally the derivation of POC value of a picture does not depend on the POC values of pictures in other layers. As IDR picture and BLA picture will force the complete POC value or the most significant bits (MSB) of the POC to be 0, the POC values of pictures within an access unit may be different (e.g. BL and EL pictures in AU(t4) in Fig. 4), which violates the constraint that the pictures in the same access unit have the same POC value. To solve this, a POC resetting process is designed for multi-layer extension, which resets the POC values for the pictures in an access unit when such pictures in different layers would get different POC values following the normal POC derivation process as specified in HEVC Version 1 standard [11]–[13]. In addition, to keep the consistency of the POC differences in reference picture set (RPS) operations, POC shifting operations are performed after resetting on previous pictures in decoding order as a decrement in POC values. The parameters for POC resetting are signalled in slice segment header extension.

### 3.3 Hybrid Coding

Unlike SVC, which has the BL coded in H.264/AVC and the BL bitstream embedded in the SVC bitstream, multi-layer extension enables the BL bitstream to be provided by external means not specified in the second version of HEVC standard. Furthermore, the BL bitstream provided by external means can be generated by any single layer encoder besides HEVC, such as H.264/AVC, MPEG-2, and etc. This feature of multi-layer extension can be referred to as hybrid codec scalability or hybrid coding. In this case, decoding of the external BL is out the scope of multi-layer extension, and hybrid coding is carried out following the decoding operations of EL specified in multi-layer extension by forwarding the reconstructed pictures after decoding external BL to EL and inserted into the EL reference lists for inter-layer prediction.

One use case for hybrid coding is long-term gradual upgrading a system by appending an HEVC EL to the existing stream coded by other standards (e.g., MPEG-2, H.264/AVC). The HEVC EL may provide higher resolution, higher dynamic



▲ Figure 4. Example multi-layer bitstream structure with layer-wise decoding.

range and/or wider color gamut to enhance the viewing quality or provide a view other than the view represented by BL to form a stereo pair for 3D viewing. Accordingly, the devices with hybrid coding will provide more vivid viewing experience, while legacy devices can still provide basic perceptual quality by discarding the HEVC EL bits. The main advantage is that a lot of bandwidth can be saved compared to simulcast solution of two separate bitstreams while maintaining backward compatibility during upgrade. However, the cost is that the devices with hybrid coding need to support a number of standards and conduct exact synchronization of HEVC EL pictures and BL pictures in both inter-layer prediction and picture output process (especially for stereo pair in 3D viewing).

### 3.4 Independent Non-Base Layer (INBL)

Multi-layer extension supports INBL. The INBL is an EL in multi-layer bitstream, which is coded without using inter-layer prediction and conforms to a single layer profile. That is, the only difference of an INBL and ordinary single layer bitstream is that the `nuh_layer_id` in NAL units in INBL stream is not equal to 0. In VPS, a flag is signalled along with profile, tier and level of a layer to indicate whether this layer is an INBL. INBL provides a simulcast layer in the multi-layer extension. This flag is also used to signal the capability of a decoder whether INBL can be processed, which is used in, for example, session negotiation. An INBL rewriting process is also designed for multi-layer extension to convert the INBL bitstream extracted from multi-layer bitstream into a bitstream conforming to a single layer profile.

## 4 Conclusions

This paper gives an overview of the concepts and HLS in multi-layer extension of HEVC Version 1 standard. Multi-layer extension is developed based on HEVC Version 1 standard and serves as a common architecture for HEVC extensions using layered structure, including SHVC and MV-HEVC in the second version of HEVC. High-level extension approach is used to multi-layer extension without changing the block level decoding operations already specified in single layer HEVC profiles. This design principle enables the implementation of SHVC and MV-HEVC to be built on existing single layer HEVC codec with additional inter-layer reference picture processing operations, which dramatically alleviates the workload of codec design. Additionally, several new designs are also developed for multi-layer extension to achieve more flexibility for applications and future extensions using layered structure. The benefits of multi-layer extension will facilitate widespread adoption of layered coding extensions of HEVC to applications.

### Acknowledgment

The authors thank the experts of ITU-T VCEG, ISO/IEC,

MPEG, the ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) and the ITU-T/ISO/IEC Joint Collaborative Team on 3D Video Coding Extension Development (JCT-VC) for their contributions.

### References

- [1] *High Efficiency Video Coding*, ITU-T Recommendation H.265 standard v1, Apr. 2013.
- [2] J. -R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012. doi: 10.1109/TCS-VT.2012.2221192.
- [3] T. K. Tan, M. Mrak, V. Baroncini, and N. Ramzan, "Report on HEVC compression performance verification testing," ITU, JCT-VC Document JCTVC-Q1011, Mar. 2014.
- [4] 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/JST-SP.2013.2283657.
- [5] *High Efficiency Video Coding*, ITU-T Recommendation H.265 standard v2, Oct. 2014.
- [6] *High Efficiency Video Coding*, ITU-T Recommendation H.265 standard v3, Apr. 2015.
- [7] R. Joshi, S. Liu, G. J. Sullivan, *et al.*, "HEVC screen content coding draft text 4," ITU, JCT-VC Document JCTVC-V1005, Oct. 2015.
- [8] K. Minoo, P. Yin, T. Lu, *et al.*, "Exploratory test model for HDR extension of HEVC," MPEG Document N15792, Oct. 2015.
- [9] *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Recommendation H.264, Jun. 2011.
- [10] L. Guo, Y. He, D.-K. Kwon, *et al.*, "TE 2: summary report on inter-layer texture prediction signaling in SHVC," ITU, JCT-VC Document JCTVC-L0022, Jan. 2013.
- [11] Y. Chen, Y.-K. Wang, and A. K. Ramasubramanian, "MV-HEVC/SHVC HLS: cross-layer POC alignment," ITU, JCT-VC Document JCTVC-N0244 & JCT-3V Document JCT3V-E0075, Jul. 2013.
- [12] G. J. Sullivan, "JCT-VC AHG report: multi-layer picture order count derivation (AHG10)," ITU, JCT-VC Document JCTVC-P0010, Jan. 2014.
- [13] Hendry Fnu, A. K. Ramasubramanian, Y.-K. Wang, *et al.*, "MV-HEVC/SHVC HLS: on picture order count," ITU, JCT-VC Document JCTVC-P0041 & JCT-3V Document JCT3V-G0031, Jan. 2014.

Manuscript received: 2015-11-10

## Biographies

**Ming Li** (li.ming42@zte.com.cn) received the BEng degree in telecommunication engineering and PhD degree in communication and information systems from Xidian University, China, in 2005 and 2010, respectively. He has been a standardization engineer in video coding in ZTE since 2010. His current research interests include video coding and multimedia communication.

**Ping Wu** (ping.wu@zte.com.cn) received the BEng degree in Electrical Engineering from Tsinghua University, Beijing, China in 1985 and received the PhD degree in signal processing from Reading University, United Kingdom in 1993. From 1993 to 1997, he was a research fellow in the area of medical data processing in Plymouth University, United Kingdom. From 1997 to 2008, he was a consultant engineer in News Digital Systems Ltd, Tandberg Television, and Ericsson. He participated in the development of ISO/IEC MPEG and ITU-T video coding standards. He also supervised the engineering team to build the High Definition H.264 encoder products for broadcasters. From 2008 to 2011, he joined Mitsubishi Electric Research Centre Europe and continued to participate in High Efficiency Video Coding (HEVC) standard development with contributions in Call for Evidence and Call for Proposal. From 2011, he has been a senior specialist in video coding in ZTE. He has many technical proposals and contributions to the international standards on video coding over past 18 years.



# SHVC, the Scalable Extensions of HEVC, and Its Applications

Yan Ye<sup>1</sup>, Yong He<sup>1</sup>, Ye-Kui Wang<sup>2</sup>, and Hendry<sup>2</sup>

(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



## SHVC, the Scalable Extensions of HEVC, and Its Applications

Yan Ye, Yong He, Ye-Kui Wang, and Hendry

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.

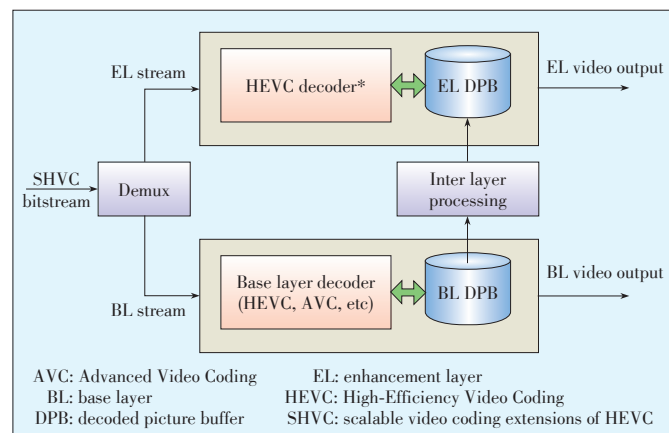
## 2 SHVC

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.

## SHVC, the Scalable Extensions of HEVC, and Its Applications

Yan Ye, Yong He, Ye-Kui Wang, and Hendry

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 port-

able, 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 PLP: physical-layer pipe		qHD: quarter high definition 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.

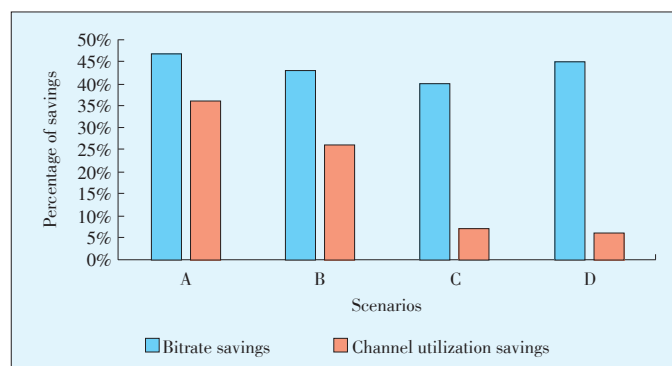
Fig. 2 shows the performance of SHVC and HEVC simulcast for each scenario. Two types of savings were calculated: 1) the percentage of average video bit rate savings, which is how much SHVC reduces the coded video bit rate compared to simulcast while maintaining the same quality (PSNR); and 2) the percentage 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 speci-



▲ Figure 2. SHVC vs. HEVC simulcast.

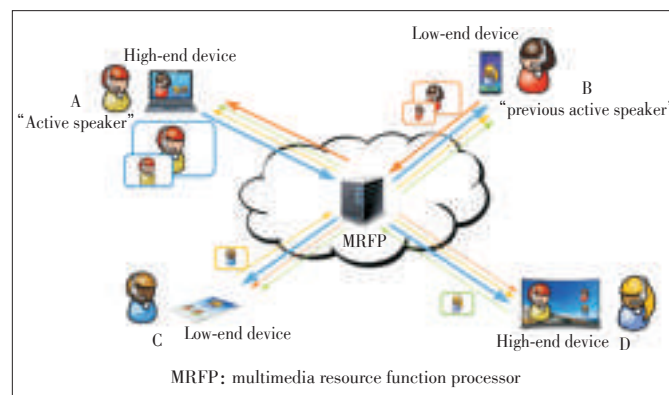
fied the media handling aspects of all 3GPP multimedia service standards, including 3GP-DASH in TS 26.247 [21], Packet-switched Streaming Service (PSS) in TS 26.234 [22], MBMS in TS 26.346 [23], Multimedia Telephony Service over IMS (MTSI) in TS 26.114 [24] (this also includes MMVC), Multimedia Messaging 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 3. The use case for MMVC and telepresence.

## SHVC, the Scalable Extensions of HEVC, and Its Applications

Yan Ye, Yong He, Ye-Kui Wang, and Hendry

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

▼ **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				

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.

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**.

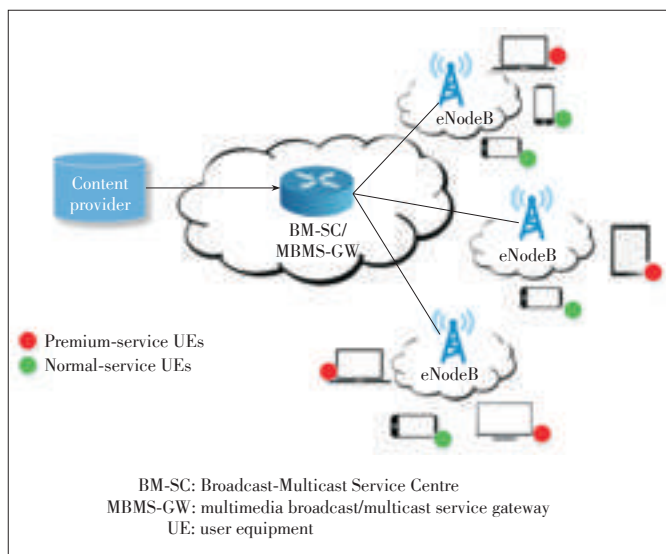
When SHVC is used in the differentiated-service MBMS use-case, the content is encoded with two layers of different spatial resolutions, and is transmitted from the content provider to the BM-SC, and all the way to the UEs. Each premium-service UE receives and decodes both layers and renders the higher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



## SHVC, the Scalable Extensions of HEVC, and Its Applications

Yan Ye, Yong He, Ye-Kui Wang, and Hendry



▲ 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

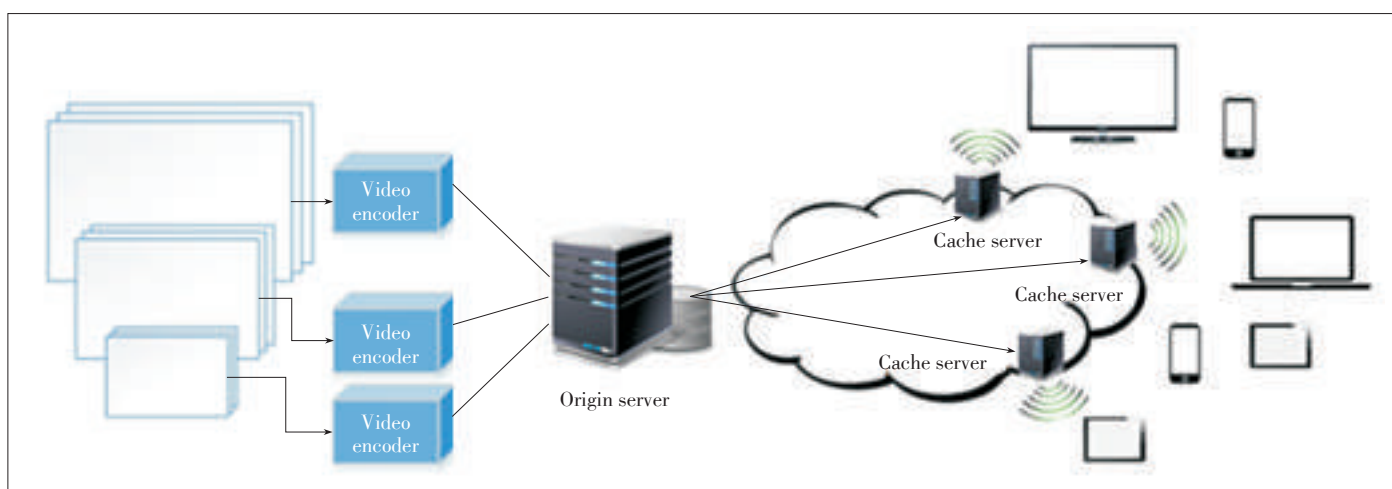
be assumed the same as when HEVC simulcast is used because UEs receiving normal-service can ignore coded data for enhancement layer. The decoding complexity overhead for UEs receiving the premium-service when SHVC is used is roughly the percentage of the number of samples in the lower resolution video 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.



## SHVC, the Scalable Extensions of HEVC, and Its Applications

Yan Ye, Yong He, Ye-Kui Wang, and Hendry

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 Conclusions

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](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, Apr. 2014.
- [11] *Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-*

## SHVC, the Scalable Extensions of HEVC, and Its Applications

Yan Ye, Yong He, Ye-Kui Wang, and Hendry

- 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. Ramasubramanian, "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.

# ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution

Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fr-hlich, and Guan-Ming Su  
(Dolby Laboratories Inc., Sunnyvale, CA 94085, USA)

## Abstract

High Dynamic Range (HDR) and Wider Colour Gamut (WCG) content represents a greater range of luminance levels and a more complete reproduction of colours found in real-world scenes. The current video distribution environments deliver Standard Dynamic Range (SDR) signal Y'CbCr. For HDR and WCG content, it is desirable to examine if such signal format still works well for compression, and to know if the overall system performance can be further improved by exploring different signal formats. In this paper, ITP (IC<sub>r</sub>C<sub>p</sub>) colour space is presented. The paper concentrates on examining the two aspects of ITP colour space: 1) ITP characteristics in terms of signal quantization at a given bit depth; 2) ITP compression performance. The analysis and simulation results show that ITP 10 bit has better properties than Y'CbCr-PQ 10bit in colour quantization, constant luminance, hue property and chroma subsampling, and it also has good compression efficiency. Therefore it is desirable to adopt ITP colour space as a new signal format for HDR/WCG video compression.

## Keywords

HDR; WCG; Y'CbCr; ITP; IC<sub>r</sub>C<sub>p</sub>

## 1 Introduction

Current video distribution environments deliver a Standard Dynamic Range (SDR) signal. For SDR content, the common practice is to apply compression on a non-constant luminance (NCL) Y'CbCr colour difference signal defined in ITU-R BT.709 [1] using a gamma transfer function (ITU-R BT.1886 [2]) and non-constant luminance 4:2:0 chroma subsampling. With the advance of display technologies, commercial interests in High Dynamic Range (HDR) and Wide Colour Gamut (WCG) content distribution are growing rapidly. Compared with conventional SDR content, HDR/WCG video content has a greater range of luminance levels and colours found in real-world scenes, and this creates a more pleasant, immersive viewing experience for people with advanced HDR displays. In order to deliver the HDR/WCG content, an HDR/WCG video distribution workflow has to be employed from content creation to final display, which comprises of post-production, encoding, transmission, decoding, colour volume mapping and display. It is desirable to have a signal format for HDR and WCG content that is not only suitable for efficient image signal encoding, but also suitable for video compression and colour volume mapping. Therefore, for

HDR/WCG content distribution, we can examine whether the conventional Y'CbCr 4:2:0 NCL signal format is still a good format to represent HDR/WCG video, and if it still compressed well by a video codec developed using SDR content. In this paper, the main focus is on compression related part (the encoding and decoding blocks) in the distribution pipeline.

MPEG is the working group formed by ISO and IEC to create standards for video and audio compression and transmission. In July 2013, MPEG started to look into the problem at the request of several studios and consumer electronics companies [3]. An Ad-Hoc Group (AhG) on HDR and WCG was established to investigate if any changes to the state-of-the-art High Efficiency Video Coding (HEVC) standard [4] are needed for HDR/ WCG video compression. For applications such as Ultra HD Blu-ray disk, it is mandatory that HDR content is transmitted using the HEVC Main 10 profile with metadata in VUI and SEI message [5]. This is commonly referred to as the "HDR-10" solution [6]. The HDR-10 essential metadata includes the signaling of the following video characteristics: SMPTE ST 2084 HDR Perceptual Quantizer transfer function (PQ-TF), ITU-R BT. 2020 [7] colour primary, and Y'CbCr non-constant luminance in ITU-R BT. 2020. In this paper, the signal format in HDR-10 is referred to as Y'CbCr-PQ 10bit. Oth-

## ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution

Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fröhlich, and Guan-Ming Su

er than defining some metadata to represent different video characteristics, “HDR-10” follows closely the common practice of SDR distribution which uses the gamma transfer function, ITU-R BT.709 colour primary, and  $Y'CbCr$  non-constant luminance in BT.709.

After several rounds of tests and demonstrations, the MPEG HDR/WCG AhG concluded that, for applications that use high bitrate compression, such as the Blu-Ray application, the performance of “HDR-10” seems to be sufficient. For applications that need compression at lower bitrates, such as broadcast and over-the-top (OTT) applications, several shortcomings of “HDR-10” were discovered, suggesting that further improvement might be necessary. In February 2015, MPEG issued a Call for Evidence (CfE) [8] to look for solutions that improve the HDR/WCG video coding performance over HEVC Main 10. A set of anchors targeting broadcast/OTT bitrates are provided using the HEVC Main 10 codec with ST 2084 [9] and ST 2086 support [10]. Anchors generated this way closely model the “HDR-10” distribution system.

As active participants of the MPEG committee work on HDR/WCG coding, Arris, Dolby, and InterDigital submitted a joint proposal (the ADI solution) [11] in response to the CfE. The joint proposal provides evidence that with a few new technologies: 1) ITP ( $IC_T C_P$ ) colour space; 2) colour enhancement filter; and 3) adaptive reshaping and transfer function, the coding performance can be further improved for HDR/WCG content [12].

This paper mainly focuses on the ITP colour space. The paper tries to answer two questions: 1) what is ITP? Compared to  $Y'CbCr$ -PQ, what is advantage of using ITP? 2) How does ITP work for compression?

The paper is organized as follows. Section 2 describes ITP colour space. Section 2.1 describes ITP conversion workflow. Section 2.2 presents ITP properties. Section 3 presents ITP compression performance for HDR and WCG video compression followed by conclusion in Section 4.

## 2 ITP Colour Space and ITP ( $IC_T C_P$ ) Colour Space

Non-Constant Luminance (NCL)  $Y'CbCr$  is the most frequently used colour space for the distribution of SDR signals.  $Y'CbCr$  is a colour difference model derived from nonlinear  $R'G'B'$  signals. For HDR signals, the ST. 2084 (also known as PQ) transfer function is applied in linear RGB space. NCL  $Y'CbCr$  has some limitations: 1) it cannot fully de-correlate intensity information from chroma information; 2) it is constrained by RGB colour primaries; therefore, the  $3 \times 3$  matrix coefficients keep on changing according to RGB colour primaries; 3) its colour difference weights are not based on perceptual model but are derived by filling a colour volume. Constant Luminance (CL)  $Y'CbCr$  was added in ITU-R BT. 2020 to adjust  $Y'CbCr$  to better de-correlate intensity from chroma. How-

ever, the conversion is significantly more complex, making it harder to use in real applications. The ITP colour space is an alternative colour space to de-correlate intensity and chroma information better matching the perceptual mechanisms of the Human Visual System (HVS) [13]. This alternate colour space is more advantageous for HDR and WCG video than the NCL  $Y'CbCr$  in signal representation. The advantage of using HVS to derive such a colour space is that the distortion introduced is perceptually minimized.

### 2.1 ITP Conversion Flow

ITP uses a colour opponent model which has similar conversion flow to NCL  $Y'CbCr$  but more closely mimics the HVS. I corresponds to brightness of the pixel nonlinearly encoded (similar to how the  $Y'$  is encoded in  $Y'CbCr$ ), T corresponds to blue-yellow colour perception and P corresponds to red-green colour perception. ITP was first introduced in 1998 and was optimized using a limited set of training data with standard dynamic range and BT. 709 colour gamuts due to the lack of HDR/WCG content at that time [14]. The proposed ITP colour space improves the original IPT by exploring higher dynamic range (up to 10000 nits) and larger colour gamuts (BT. 2020). Considering the non-trivial changes over the original IPT and to follow the  $Y'CbCr$  practice to have blue-related colour component prior to red-related colour component, the new name ITP or  $IC_T C_P$  colour space is adopted to refer to this variation of the original IPT colour space.

To better understand ITP, the early stages of human colour vision are described as follows (Fig. 1) [13]:

- 1) Incoming light strikes the three photo receptors (cones) in the eye that have their peak sensitivity in the (L)ong, (M)edium, and (S)hort wavelengths;
- 2) This linear light is transduced (converted) into a non-linear signal response to reduce dynamic range;
- 3) The non-linear output goes through a colour differencing process to extract important information and separates the signal into three distinct pathways;
- 4) The brain sees three colour opponent channels.

ITP conversion steps (Fig. 2) are as follows:

- 1) Compute LMS response;
- 2) Apply non-linear encoding PQ;
- 3) Apply colour differencing equation.

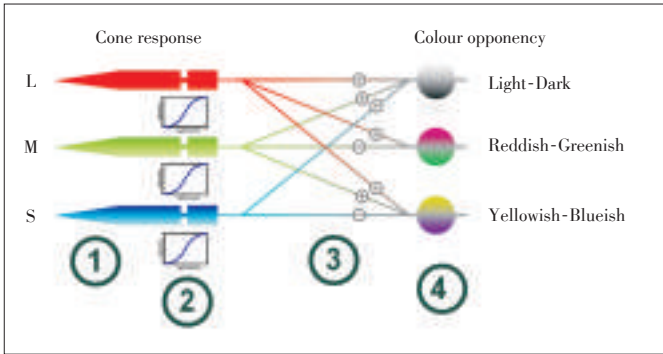
The similar complexity of the conversion as  $Y'CbCr$  allows mass deployment in a wide range of devices. Essentially the existing devices can just change the  $3 \times 3$  matrix.

The conversion matrix from the CIE XYZ tri-stimulus values to LMS (derived  $RGB_{2020}$  to LMS) and LMS to ITP are listed below. Equ. (1) shows the Conversion from XYZ to LMS colour space, (2) shows the Conversion from  $RGB_{2020}$  to LMS colour space, and (3) shows the Conversion from  $L'M'S'$  to ITP colour space. Note that the coefficients in the conversion matrices shown in this paper are the rounded decimal representation of the real conversion matrices. They may have higher precision

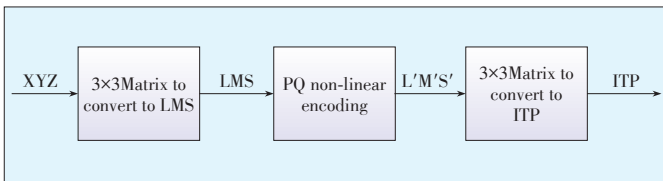


## ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution

Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fröhlich, and Guan-Ming Su



▲ Figure 1. Opponent colour model in HVS.



▲ Figure 2. XYZ to ITP conversion.

or fixed point representation depending on implementation needs.

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = \begin{pmatrix} 0.3592 & 0.6976 & -0.0358 \\ -0.1922 & 1.1004 & 0.0755 \\ 0.0070 & 0.0749 & 0.8434 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = \begin{pmatrix} 0.4120 & 0.5239 & 0.0641 \\ 0.1667 & 0.7204 & 0.1129 \\ 0.0241 & 0.0755 & 0.9004 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} I \\ P \\ T \end{pmatrix} = \begin{pmatrix} 0.5000 & 0.5000 & 0.0000 \\ 1.6137 & -3.3234 & 1.7097 \\ 4.3780 & -4.2455 & -0.1325 \end{pmatrix} \begin{pmatrix} L' \\ M' \\ S' \end{pmatrix} \quad (3)$$

### 2.2 ITP Properties

When designing a colour space, the main goal is to minimize colour distortion and prevent visible quantization artifacts when images are represented with a given number of digital codewords (i.e., given bit depth). Another requirement is to decorrelate the chroma information from luma information to enable colour subsampling, which is important for video compression. In the context of HDR and WCG, and due to various displays in market which supports different dynamic range and colour gamut, a colour space should fit for colour volume mapping as well. In the following, a set of psychophysical experiments have been conducted to validate the advantages of ITP over Y'CbCr-PQ for HDR and wide-gamut imaging. Compared with Y'CbCr-PQ, ITP has the following properties:

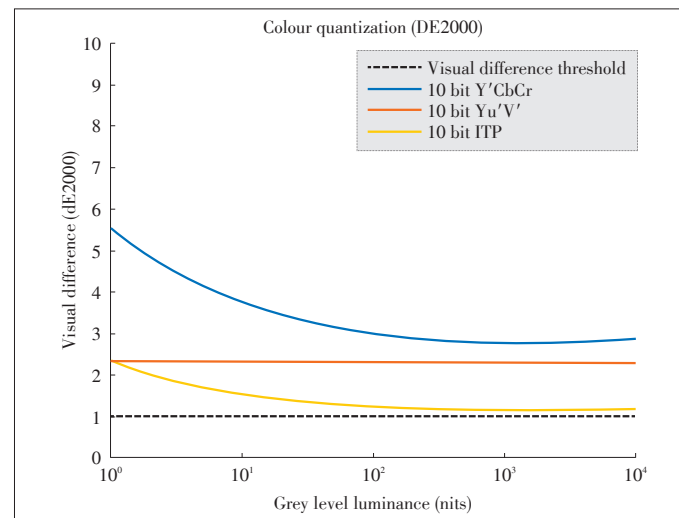
- 1) Better signal representation (smaller just-noticeable-difference) in colour quantization in 10 bits;
- 2) Improved intensity prediction (constant luminance);
- 3) Better predicted lines of constant hue for worst case;
- 4) Friendliness to 4:2:0 chroma downsampling.

#### 2.2.1 Baseband Property in Colour Quantization

Baseband signal encoding refers to the representation of a linear light HDR signal in integer codewords in a given bit depth. Ideally, the higher the bit depth, the easier the quantized signal can preserve the dynamics in the original linear light signal. However, due to practical considerations, the highest existing pipeline for broadcast is limited to 10 bits. So it is very important to investigate how good ITP 10 bit baseband property can be. It decides how good the signal can start with and thus impact the full chain performance of HDR and WCG content distribution. Mathematical computation shows that ITP has the best overall baseband performance compared with Y'CbCr and Yu'v' when quantized to 10 bits (Fig. 3). The industry accepted DE2000 metric is used to measure the visual difference. If the value is below the detection threshold of one “just noticeable difference” JND, no noticeable colour quantization artifact can be observed. The value of dE2000 for Y'CbCr-PQ 10b is between 3.0 and 5.5. For Yu'v'-PQ, it is 2.3. For ITP, it is about 1.0 above 100 nits which is the JND threshold. The better colour quantization property of ITP is due to the fact that ITP is more perceptually uniform than the other colour spaces [15].

#### 2.2.2 Constant Luminance Property

Constant luminance encoding is more effective in reducing crosstalk between luma and chroma components than the conventional NCL encoding method. Therefore, a colour space which has better constant luminance property tends to have better chroma downsampling, such as 4:2:0. Both subjective experiment and theory shows that ITP outperforms NCL Y'CbCr in intensity prediction. In the subjective experiment, 11 participants matched the intensity of a colour patch with a reference



▲ Figure 3. Visual difference of colour space.



## ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution

Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fröhlich, and Guan-Ming Su

neutral. The data gathered was used to test various colour spaces. ITP outperforms NCL Y'CbCr in intensity prediction (indicated by the higher correlation to the reference) (**Fig. 4**). The property is also validated in scientific analysis, where uniformly distributed RGB samples are generated and is converted in ITP, NCL Y'CbCr and CL Y'CbCr, and the correspondence is shown in **Fig. 5**. The I in ITP correlates better with constant luminance Y' than the NCL Y' from Y'CbCr.

### 2.2.3 Hue Property

For colour volume mapping, since observers perceiving changes in hue is more impactful than changes in lightness or chroma, it is desirable to have a colour space as hue-linear as possible. Linear hue lines make it very easy to model the mapping process with the hue-preservation requirement. ITP was designed for linear hue property, so it has better hue linearity. A psychophysical experiment was conducted to determine lines of constant hue at multiple hue angles (**Fig. 6**). ITP more closely follows the lines of predicted constant hue than Y'CbCr for the worst case measured by the maximum absolute hue deviation. The most notable improvement with hue linearity is the lack of large deviations in ITP as opposed to those found on the right of the constant hue Y'CbCr plot. **Table 1** showed the average and maximum absolute hue deviation of ITP and Y'

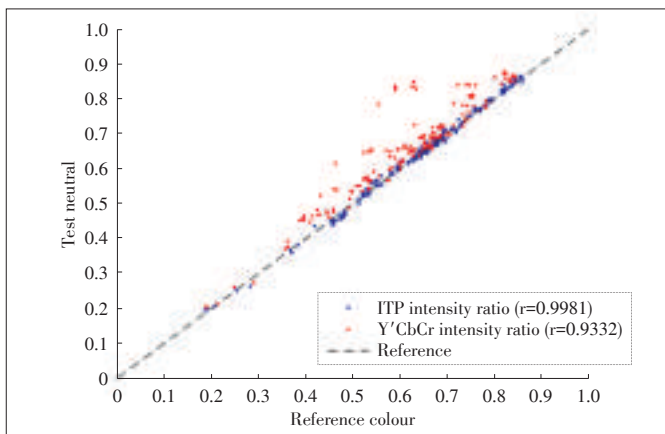
CbCr, respectively.

### 2.2.4 Chroma Downsampling Property

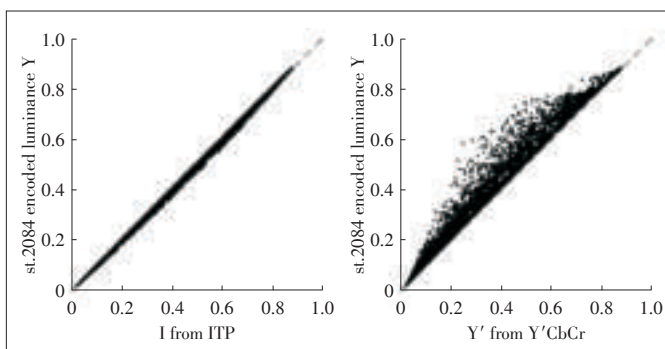
Similar to Y'CbCr, ITP is also friendly to 4:2:0 chroma downsampling, which is the common chroma sampling format used in video compression. **Table 2** shows the difference in dB in objective metrics computed from a conversion only workflow for ITP and Y'CbCr-PQ. The details of those objective metrics are referred to in [12]. The MPEG CfE chroma down/up sampling filters are applied. The conversion flow is as follows: RGB 4:4:4 (12 or 16 bit depending on content) → Y'CbCr-PQ 4:2:0 10 bit / ITP 4:2:0 10 bit → RGB 4:4:4 (original bitdepth). ITP has overall higher PSNR in luminance channel (Y) and overall colour metrics (DE) than Y'CbCr NCL, and is suitable for compression in the 4:2:0 domain.

## 3 ITP Compression Performance

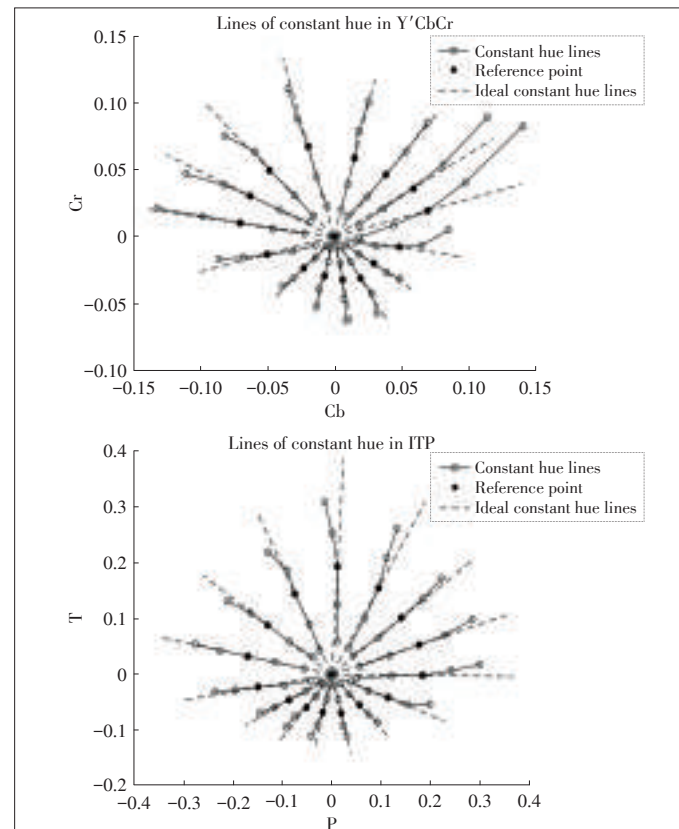
As well as having better baseband signal properties than the



▲ **Figure 4.** Iso-intensity performance.



▲ **Figure 5.** Comparison of constant luminance performance of ITP and NCL Y'CbCr-PQ.



▲ **Figure 6.** Constant hue.

▼ **Table 1.** Absolute hue deviation of ITP and Y'CbCr

	Absolute Hue Deviation (degrees)	
	Average	Maximum
ITP	2.34	7.79
Y'CbCr	2.95	16.6

## ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution

Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fröhlich, and Guan-Ming Su

▼ **Table 2. Objective metric differences for conversion-only (ITP 10bit - Y'CbCr-PQ 10bit)**

Sequence	Diff tPSNR Y	Diff DEPSNR
FireEater	6.57	0.97
Tibul2	10.8	0.83
Market3	7.81	0.17
AutoWelding	8.31	0.17
BikeSparklers	5.18	0.04
ShowGirl2	4.9	0.15
MagicHour	2.54	0.05
WarmNight	6	0.06
BalloonFestival	7.36	0.09
<b>Average (dB)</b>	<b>6.61</b>	<b>0.28</b>

NCL Y'CbCr-PQ, ITP can compress well. To evaluate the compression performance of ITP, two studies have been conducted: 1) compare transform coding gain with KLT; 2) compare covariance of ITP with Y'CbCr. We formed a test set of 19 frames representing all the scenes from MPEG HDR/WCG AhG sequences, i.e., one representative frame from each scene.

Transform coding gain is one metric to measure compression performance. It is defined by the ratio of the arithmetic mean to the geometric mean of the variances of the variables in the new coordinates, properly scaled by the norms of the synthesis basis functions for nonunitary transforms [16]. The coding gain is usually measured in decibels, representing the reduction in quantization noise by quantizing in the transformed domain instead of the original domain [16]. The test shows that the coding gain of ITP is 12.42 dB, while the coding gain using optimal KLT is 13.16 dB.

Another important statistical indicator for compression is the covariance matrix of the signal for 3 channels. We computed the covariance matrix of the test set with BT. 2020 Y'CbCr-PQ and ITP, both in 10-bit 4:4:4 Standard Range.

The covariance of 3 channels for Y'CbCr-PQ 10bit case is shown in **Table 3**, and that for ITP 10bit case is shown in **Table 4**.

By comparing the above two covariance matrix, we found that the variance of P channel is about four times of Cb channel and the variance of T channel is about four times of Cr channel, respectively. The cross-variances of IP and IT are about twice that of YCb and YCr, respectively. The cross-variance of PT is about four times of CbCr. This indicates that if we reduce the signal of P/T by half, i.e., representing P/T with 9 bit, the covariance matrix should be close to Y'CbCr case. **Table 5** is the covariance matrix for the newly generated ITP signal. They are indeed very close to Y'CbCr case in terms of covariance.

The conversion only results for I 10bit and PT 9bit compared with Y'CbCr 10bit case is listed in **Table 6**. By comparing Table 6 with Table 2, the conversion-only benefit of ITP

over Y'CbCr is reduced but is still retained for majority testing content. One exception is the sequence BalloonFestival featuring very saturated colours, suggesting 9 bit is not good enough to signal chroma components for this content.

The compression simulation was also performed to test the performance of ITP. ITP colour space is implemented in HDRTools 0.8.1 [17]. HM16.2 [18] is used for compression test. The test sequences and targeted bitrate is listed in **Table 7**. For I 10bit and TP 9bit case, we used the same fixed QP as in the Y'CbCr - PQ anchor case. The compression results showed similar bit rate as the anchor. The BD-rate calculated using the suggested metrics is shown in **Table 8**. This is a fair comparison with Y'CbCr because a fixed scalar is used for P and T. Alternatively, we can simply encode the signal with HEVC by setting luma bit depth to 10 and chroma bit depth to 9. This shows that for DE100, which is considered as a performance indicator for colour reproduction, ITP with static reshap-

▼ **Table 3. The covariance of 3 channels for Y'CbCr-PQ 10bit case**

	Y	Cb	Cr
Y	2.8949	-0.1304	0.0766
Cb	-0.1304	0.0730	-0.0314
Cr	0.0766	-0.0314	0.0321

▼ **Table 4. The covariance of 3 channels for ITP 10bit case**

	I	T	P
I	2.5430	-0.2146	0.2099
T	-0.2146	0.2658	-0.1344
P	0.2099	-0.1344	0.1912

▼ **Table 5. The covariance matrix for the newly generated ITP signal**

	I	T	P
I	2.8653	-0.1233	0.1214
T	-0.1233	0.0783	-0.0396
P	0.1214	-0.0396	0.0564

▼ **Table 6. Objective metric differences for conversion-only (ITP I 10bit PT 9bit - Y'CbCr-PQ 10bit)**

Sequence	Diff tPSNR Y	Diff DEPSNR
FireEater	5.27	0.73
Tibul2	9.51	0.52
Market3	6.72	0.02
AutoWelding	7.74	0.13
BikeSparklers	4.84	0.02
ShowGirl2	4.23	0.05
MagicHour	2.17	0.04
WarmNight	5.58	0.04
BalloonFestival	6.39	-0.22
<b>Average (dB)</b>	<b>5.83</b>	<b>0.15</b>

## ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution

Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fröhlich, and Guan-Ming Su

▼ Table 7. HDR/WCG test sequences and target rate points (kbps)

Class	Seq	Sequence name	Rate 1	Rate 2	Rate 3	Rate 4
A	S00	FireEater2Clip4000r1	1922	1260	812	521
	S01	Tibul2Clip4000r1	6101	2503	970	403
	S02	Market3Clip4000r2	7913	4224	2311	1248
B	S03	AutoWeldingClip4000	3157	1383	778	454
	S04	BikeSparklersClip4000	6119	4085	2184	1261
C	S05	ShowGirl2TeaserClip4000	3316	1652	971	574
D	S06	StEM_MagicHour	3959	2205	1302	771
	S07	StEM_WarmNight	2441	1328	780	462
G	S08	BalloonFestival	6644	3767	2156	1276

▼ Table 8. Compression results (BD rates) compared to Y'CbCr-10b for I10b TP9b

	X	Y	Z	XYZ	<sup>t</sup> OSNR-XYZ	DE100	MD100	PSNRL100
FireEaterClip4000r1	-25.0%	-11.4%	53.4%	2.0%	-3.3%	-23.5%	-19.5%	-11.1%
Market3Clip4000r2	-6.1%	-1.4%	-3.8%	-3.8%	-5.5%	-23.5%	-7.2%	-1.7%
Tibul2Clip4000r1	-21.9%	-13.5%	150.3%	5.4%	11.5%	-15.6%	-7.2%	-11.4%
AutoWelding	-10.6%	0.2%	13.0%	1.5%	1.8%	-19.0%	-20.3%	3.1%
BikeSparklers	-10.5%	-1.0%	4.9%	-1.5%	-3.3%	-21.2%	-16.8%	0.0%
ShowGirl2Teaser	-10.9%	-1.7%	-7.7%	-6.7%	-10.4%	-22.6%	-25.8%	-2.2%
StEM_MagicHour	-9.6%	-1.2%	-3.9%	-4.7%	-5.2%	-15.9%	-19.7%	-1.0%
StEM_WarmNight	-13.0%	-1.0%	6.4%	-0.8%	-1.7%	-22.1%	-28.2%	-0.5%
BalloonFestival	-7.8%	-2.2%	2.6%	-1.5%	-2.8%	-7.7%	-23.5%	-2.8%
<b>Overall</b>	<b>-12.8%</b>	<b>-3.7%</b>	<b>23.9%</b>	<b>-1.1%</b>	<b>-2.1%</b>	<b>-19.0%</b>	<b>-18.7%</b>	<b>-3.1%</b>

ing can gain 19% over Y'CbCr-PQ.

These findings suggest that ITP 10 bit signal contains more colour information than Y'CbCr-PQ 10 bit signal. Since the baseband signal has much better representation in colour, it gives compression much more flexibility for having a “better” signal to start with. Considering this aspect, a technology called adaptive reshaping is incorporated into ITP to adaptively adjust the quantization of luma and chroma components and maximize coding efficiency. The evidence is shown in the MPEG CfE ADI proposal [11] and CfE test results report [19], and MPEG Core Experiment CE2.1.1 results [20] where advanced reshaping is applied in ITP colour space. In all those tests, ITP has shown superior compression performance compared to the MPEG CfE anchor. **Table 9** list results in MPEG HDR/WCG CE2.1.1.

When compared on an HDR display, the ITP with advanced reshaping can significantly improve compression performance of the HEVC Main 10 Anchors. The colour patches/blotches are mitigated substantially in low to medium bitrate compression. Besides, it can also improve texture preservation. **Fig. 7** shows the snapshots taken during the side-by-side (SbS) viewing on the HDR reference display Pulsar, for the test sequence

▼ Table 9. Compression results (BD rates) of MPEG HDR/WCG CE2.1.1

	X	Y	Z	XYZ	<sup>t</sup> OSNR-XYZ	DE100	MD100	PSNRL100
FireEaterClip4000r1	-11.8%	4.9%	19.2%	2.4%	-0.3%	-21.3%	-32.3%	-6.4%
Tibul2Clip4000r1	-8.7%	2.9%	11.6%	-0.7%	-4.8%	-22.1%	-16.0%	-5.3%
Market3Clip4000r2	12.4%	20.6%	-13.2%	3.6%	-12.1%	-70.5%	0.0%	-17.2%
AutoWelding	-15.4%	-0.8%	-13.6%	-10.6%	-10.9%	-48.0%	-21.8%	1.8%
BikeSparklers	-16.8%	-5.1%	-17.2%	-13.8%	-16.5%	-48.1%	-11.6%	-5.0%
ShowGirl2Teaser	4.9%	19.3%	-6.6%	4.7%	1.5%	-48.6%	0.0%	-5.4%
StEM_MagicHour	-8.2%	2.5%	-10.3%	-6.8%	-8.5%	-34.5%	-21.4%	1.3%
StEM_WarmNight	-6.0%	9.5%	-5.0%	-1.5%	-2.7%	-42.7%	-45.9%	2.0%
BalloonFestival	56.9%	86.9%	110.2%	88.5%	28.7%	-45.0%	-77.9%	-1.5%
<b>Overall</b>	<b>0.8%</b>	<b>15.6%</b>	<b>8.3%</b>	<b>7.3%</b>	<b>-2.8%</b>	<b>-42.3%</b>	<b>-25.2%</b>	<b>-4.0%</b>



▲ Figure 7. Market3 (from Technicolor) coded at R3: (a) ADI (2305 kbps), (b) Anchor (2311 kbps).

es “Market3” (copyright @ Technicolor) in MPEG CfE ADI solution. The circled areas show the most significant improvements over the anchor. For example, at similar bitrates, the details on the wall and wood frames in Market3 are better preserved (Fig. 7).

## 4 Conclusions

In this paper, ITP 10 bit is shown to have better baseband properties than Y'CbCr-PQ 10 bit. The compression performance of ITP 10 bit is also justified with compression results both in MPEG HDR/WCG CfE and following Core Experiments. The other property of ITP also shows that it is a good fit for colour volume mapping too. ITP is shown to work well for full HDR and WCG video delivery pipeline. Therefore, it is desirable to endorse ITP as a new signal format for HDR/WCG signal.

## Acknowledgment

We would like to acknowledge Y. He, Y. Ye, and L. Kerofsky from InterDigital and D. Baylon, Z. Gu, A. Luthra, K. Minoo from Arris to work together for the joint CfE contribution.

## References

- [1] *Parameter Values for the HDTV Standards for Production and International Programme Exchange*, ITU-R BT.709, 2015.

## ITP Colour Space and Its Compression Performance for High Dynamic Range and Wide Colour Gamut Video Distribution

Taoran Lu, Fangjun Pu, Peng Yin, Tao Chen, Walt Husak, Jaclyn Pytlarz, Robin Atkins, Jan Fröhlich, and Guan-Ming Su

- [2] *Reference Electro-Optical Transfer Function for Flat Panel Displays Used in HDTV Studio Production*, ITU-R BT.1886, 2011.
- [3] H. Basse, W. Aylsworth, S. Stephens, *et al.*, "Proposed standardization of XYZ image," Doc. m30167, Vienna, Austria, Jul. 2013.
- [4] *HEVC*, Recommendation ITU-T H.265, International Standard ISO/IEC 23008-2, 2013.
- [5] Blu-ray Disc Association, "Ultra HD Blu-ray Video Parameters Liaison Information," Doc. m36740, Warsaw, Poland, Jun. 2015.
- [6] D. Le Gall, A. Tourapis, M. Raulet, *et al.*, "High dynamic range with HEVC main10," JCTVC-U0045, Warsaw, Poland, Jun. 2015.
- [7] *Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange*, ITU-R BT.2020, 2015.
- [8] A. Luthra, E. Francois, and W. Husak, "Call for evidence (CfE) for HDR and WCG video coding," Doc. N15083, Geneva, Switzerland, Feb. 2015.
- [9] *Electro-Optical Transfer Function for High Dynamic Range Reference Display*, Society of Motion Picture and Television Engineers ST 2084, 2014.
- [10] *Electro- Mastering Display Colour Volume Metadata Supporting High Luminance and Wide Colour Gamut Images*, Society of Motion Picture and Television Engineers ST 2086, 2014.
- [11] D. Baylon, Z. Gu, A. Luthra, *et al.*, "Response to call for evidence for HDR and WCG video coding: Arris, Dolby and InterDigital," Doc. m36264, Warsaw, Poland, Jul. 2015.
- [12] A. Luthra, E. Francois, and W. Husak, "Call for Evidence (CfE) for HDR and WCG Video Coding," Doc. N15083, Geneva, Switzerland, Feb. 2015.
- [13] G. M. Johnson, X. Song, E. D. Montag, and M. D. Fairchild, "Derivation of a colour space for image colour difference measurement," *Colour Research & Application*, vol. 35, no. 6, pp. 387–400, 2010.
- [14] F. Ebner and M. D. Fairchild, "Development and testing of a colour space (IPT) with improved hue uniformity," in *Colour and Imaging Conference*, Scottsdale, USA, Nov. 1998, pp. 8–13.
- [15] J. Froehlich, T. Kunkel, R. Atkins, *et al.*, "Encoding colour difference signals for high dynamic range and wide gamut imagery," in *Colour and Imaging Conference*, Darmstadt, Germany, Oct. 2015, pp. 240–247.
- [16] H. Malvar, G. Sullivan, and S. Srinivasan, "Lifting-based reversible colour transformations for image compression," in *Proc. SPIE 7073, Application of Digital Image Processing XXXI*, San Diego, USA, 2008. doi: 10.1117/12.797091.
- [17] MPEG SVN [Online]. Available: <http://wg11.sc29.org/svn/repos/Explorations/XYZ/HDRTools/tags/0.8.1>
- [18] HEVC [Online]. Available: [https://hevc.hhi.fraunhofer.de/svn/svn\\_HEVCSoftware/tags/HM-16.2](https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.2)
- [19] MPEG requirement, "Test Results of Call for Evidence (CfE) for HDR and WCG Video Coding", Doc. N15350, July 2015, Warsaw, Poland.
- [20] D. Baylon, Z. Gu, A. Luthra, *et al.*, "CE2.1.1: single layer HDR-only solution based on m36264", Doc. m37070, Geneva, Switzerland, Oct. 2015.

Manuscript received: 2015-11-28

## Biographies

**Taoran Lu** (tlu@dolby.com) is currently a Staff Researcher with Dolby Laboratories, Inc. She got her PhD in electrical and computer engineering from the University of Florida in December 2010 and joined Dolby in January 2011. Her research interest is on image/video processing and compression, especially on high dynamic range video distribution. She has been an active participant in the MPEG/ITU-T video compression standardization expert group. She has authored and co-authored many academic journal and conference papers, standard contributions and US patents.

**Fangjun Pu** (Fangjun.Pu@dolby.com) is currently a Research Engineer at Image Technology Department in Dolby Laboratories, Inc. She got her Master degree in Electrical Engineering Department from University of Southern California in May 2014. Her research interest is about image/video processing and compression. She is an active participant in MPEG video compression HDR/WCG related standardizations. She has authored or co-authored several standard contributions and conference papers.

**Peng Yin** (pyin@dolby.com) is Senior Staff Researcher at Image Technology Department at Dolby Laboratories, Inc. from 2010. Before joining Dolby Laboratories, she worked at Corporate Research, Thomson Inc./Technicolor. She received her PhD degree from Princeton University in 2002. Her research interest is video processing

and compression. She is very active in MPEG/VCEG video coding related standardizations and has many publications and patents. She received the IEEE Circuits and Systems Society Best Paper Award in 2003.

**Tao Chen** (tchen@dolby.com) holds a PhD degree in computer science. Since 2011, he has been with Dolby Labs as Director of Applied Research. Prior to that, he was with Panasonic Hollywood Lab in Universal City, CA and Sarnoff Corporation in Princeton, NJ. His research interests include image and video compression, 3D video processing and system, and HDR video technology. Dr. Chen has served as session chairs and has been on technical committees for a number of international conferences. He was appointed vice chair of a technical group for video codec evaluation in the Blu-ray Disc Association in 2009. Dr. Chen was a recipient of an Emmy Engineering Award in 2008. He received Silver Awards from the Panasonic Technology Symposium in 2004 and 2009. In 2002, he received the Most Outstanding Ph.D. Thesis Award from the Computer Science Association of Australia and a Mollie Holman Doctoral Medal from Monash University.

**Walt Husak** (WJH@dolby.com) is the Director of Image Technologies at Dolby Labs, Inc. He began his television career at the Advanced Television Test Center (ATTC) in 1990 carrying out video objective measurements and RF multipath testing of HDTV systems proposed for the ATSC standard. Joining Dolby in 2000, Walt has spent his early years studying and reporting on advanced compression systems for Digital Cinema, Digital Television, and Blu-ray. He has managed or executed visual quality tests for DCI, ATSC, Dolby, and MPEG. He is now a member of the CTO's office focusing his efforts on High Dynamic Range for Digital Cinema and Digital Television. Walt has authored numerous articles and papers for a number of major industry publications. Walt is an active member of SMPTE, MPEG, JPEG, ITU-T, and SPIE.

**Jaclyn Pytlarz** (Jaclyn.Pytlarz@dolby.com) holds a BS degree in Motion Picture Science from Rochester Institute of Technology. She has worked at Dolby Laboratories since 2014 as an Engineer in the Applied Vision Science Group inside Dolby's Advanced Technology Group. Prior to work at Dolby, she worked at the Academy of Motion Picture Arts and Sciences as an Imaging Science Intern and iCONN Video Production in 2013 and 2012 accordingly. Her main areas of research include vision and color science as it relates to developing technology for high dynamic range and wide color gamut displays as well signal processing for future compatibility.

**Robin Atkins** (Robin.Atkins@dolby.com) has degrees in Electrical Engineering and Engineering Physics. His career in color and imaging science began while designing High Dynamic Range displays at Brightside Technologies. These displays revealed a fascinating host of new challenges in color appearance, which he is now working to address as part of the Applied Vision Science Group at Dolby Labs. His main focus is on building color management systems for mapping High Dynamic Range and Wide Color Gamut content to a wide range of consumer display devices, and solving the question of how to best represent large color volumes for content distribution.

**Jan Fröhlich** (jfree@dolby.com) is PhD-Student at the University of Stuttgart. He is currently working on high dynamic range and wide color gamut imaging and gamut mapping. Fröhlich contributed to multiple research projects on new acquisition, production and archiving systems for television and cinema and has been involved in a number of technically groundbreaking film projects, such as Europe's first animated stereoscopic feature film and the HdM-HDR-2014 high dynamic range & wide gamut video dataset. Before starting the PhD he was Technical Director at CinePostproduction GmbH in Germany. He is member of SMPTE, IS&T, SPIE, FK-TG, and the German Society of Cinematographers (BVK).

**Guan-Ming Su** (guanming.su@dolby.com) is with Dolby Labs, Sunnyvale, CA. Prior to this he has been with the R&D Department, Qualcomm, Inc., San Diego, CA; ESS Technology, Fremont, CA; and Marvell Semiconductor, Inc., Santa Clara, CA. He is the inventor of 50+ U.S. patents and pending applications. He is the co-author of 3D Visual Communications (John Wiley & Sons, 2013). He served as an associate editor of *Journal of Communications*; and Director of review board and R-Letter in IEEE Multimedia Communications Technical Committee. He also serves as the Technical Program Track Co-Chair in ICCCN 2011, Theme Chair in ICME 2013, TPC Co-Chair in ICNC 2013, TPC Chair in ICNC 2014, Demo Chair in SMC 2014, General Chair in ICNC 2015, and Area Co-Chair for Multimedia Applications in ISM 2015. He is the Executive Director of Industrial Governance Board in Asia Pacific Signal and Information Processing Association (APSIPA) since 2014. He is a senior member of IEEE. He obtained his Ph.D. degree from University of Maryland, College Park.



# DASH and MMT and Their Applications in ATSC 3.0

Yiling Xu<sup>1</sup>, Shaowei Xie<sup>1</sup>, Hao Chen<sup>1</sup>, Le Yang<sup>2</sup>, and Jun Sun<sup>1</sup>

(1. Shanghai Jiaotong University, Shanghai 200000, China;

2. Jiangnan University, Wuxi 214000, China)

## Abstract

Despite the success of MPEG-2 Transport Stream (TS) being used to deliver services in broadcast channels, the increase of on-demand viewing of multimedia content over IP with browser-centric media endpoints introduces a new requirement for more individualized and flexible access to content. This has resulted in alternatives to MPEG-2 TS. While the needs of interactive broadcast services (such as personalized advertisement or selection of audio stream with a language suitable for a specific user) grow there is an active standardization work under going for the next generation broadcasting systems. To best enable a complete system of hybrid broadcast and broadband services, Advanced Television Systems Committee (ATSC) 3.0 has developed an enhanced broadcast transport method named Real-Time Object Delivery over Unidirectional Transport (ROUTE)/DASH for delivery of DASH-formatted content and non-real time (NRT) data. Additionally, for broadcasting, ATSC 3.0 has also adopted MPEG Media Transport (MMT) standard, which inherits major advantageous features of MPEG-2 TS and is very useful in real-time streaming delivery via a unidirectional delivery network. This paper mainly describes features and design considerations of ATSC 3.0, and discusses the applications of the transport protocols used for broadcasting, i.e., ROUTE/DASH and MMT, whose comparative introductions are also presented in details.

## Keywords

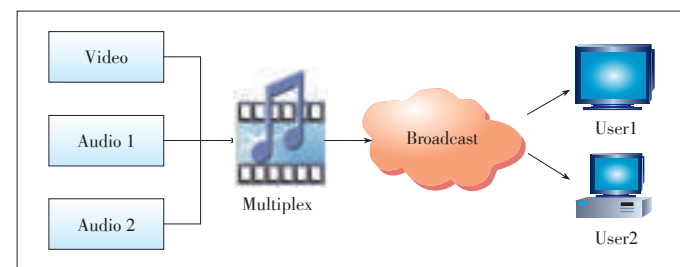
ATSC 3.0; ROUTE/DASH; MMT; Hybrid delivery; next-generation broadcasting system

## 1 Introduction

The rapid increase in the volume of multimedia content, together with the needs of flexible and interactive multimedia broadcast services such as personalized advertisement, has changed significantly the multimedia service environments. To address the associated emerging challenges [1] for multimedia delivery, standardization has been taken for developing next generation broadcasting systems that are independent content delivery systems while being able to exploit broadband networks more frequently. The resultant Advanced Television Systems Committee (ATSC) 3.0 standard incorporates a number of innovative features not typical in a conventional broadcasting system, and it is the very first full IP-based broadcasting standard. ATSC 3.0 assumes hybrid systems [2] with the media delivery channel consisting of broadcast and broadband networks.

Prior to ATSC 3.0, the most successful standards for multimedia content delivery are MPEG-2 [3], proposed by MPEG and the associated ISO Base Media File Format (ISO BMFF) [4]. Specifically, the MPEG-2 Transport Stream (TS) [5] provides very efficient methods of multiplexing multiple audiovisual

data streams into one delivery stream according to consumption order, (Fig. 1). This makes MPEG-2 TS an ideal solution for broadcasting multimedia contents in cases where a large number of users request the same content. However, MPEG-2 TS may be inadequate to fulfill the emerging needs for more individualized and flexible access to the content brought by e.g., personalized on-demand viewing of multimedia contents. This is also the case for the ISO BMFF, which stores metadata containing information for synchronized playback separately from the compressed media data. In particular, it is difficult to access a certain portion of the ISO BMFF content, e.g., to locate and retrieve an audio stream with a specific



▲ Figure 1. Multiplex of multiple audiovisual streams in MPEG-2 TS.



**DASH and MMT and Their Applications in ATSC 3.0**  
Yiling Xu, Shaowei Xie, Hao Chen, Le Yang, and Jun Sun

language during playback.

As a solution to immediate market demands, HTTP adaptive streaming [6], [7] is currently being used to deliver all broadband streaming IP content. Meanwhile, recognizing the drawbacks of existing standards and systems, MPEG developed the Dynamic Adaptive Streaming over HTTP (DASH) [8] standard by converging different HTTP adaptive streaming technologies with a focus on the adaptive streaming of media content over the legacy HTTP delivery environment [9]. Although HTTP streaming over TCP is suitable for broadband/unicast delivery, it is not an appropriate end-to-end delivery mechanism for broadcasting. To better support hybrid broadcast and broadband services, an enhanced broadcast transport method, referred to as Real-Time Object Delivery over Unidirectional Transport (ROUTE) [10]/DASH, has been proposed for delivering DASH-format content and non-real time (NRT) data over broadcast channels.

MPEG-DASH has been successful in commercial online video markets [11], [12], and MPEG turned its attention to address the new challenges of multimedia content delivery in Internet delivery environments beyond HTTP. The MPEG Media Transport (MMT) standard [13] has been recently proposed as part of the ISO/IEC 23008 High Efficiency Coding and Media Delivery in Heterogeneous Environments (MPEG-H) standard suite [14], [15]. It assumes IP networks with in-network intelligent caches close to the receiving entities. The caches actively pre-fetch the content and also adaptively packetize and push cached content to receiving entities. MMT also assumes a network environment where content can be accessed at a finer grain with unique identifiers regardless of the specific service providers and their locations [16].

Recognizing the benefits and potential for flexible and interactive multimedia services in hybrid networks, ATSC 3.0 uses both ROUTE/DASH and MMT standards. In the remainder of this paper, we give a comparative review of DASH and MMT systems with introduction of their applications in ATSC 3.0. The discussion starts in Section 2 with the design considerations of ATSC 3.0 for the next generation broadcast and the employed signaling structure. The ROUTE/DASH and MMTP/MPU are presented in the Section 3. In Section 4, content delivery in broadcast is described in detail, including the media encapsulation, the delivery of streaming services and the NRT content, as well as the system models for content delivery. Section 5 introduces the hybrid delivery modes with ROUTE/DASH and MMTP/MPU used in broadcast. Section 6 concludes the paper.

**2 ATSC 3.0 Overview**

**2.1 Receiver Protocol Stack**

To address the need for a unified receiver protocol stack that can handle the diverse potential service types and delivery

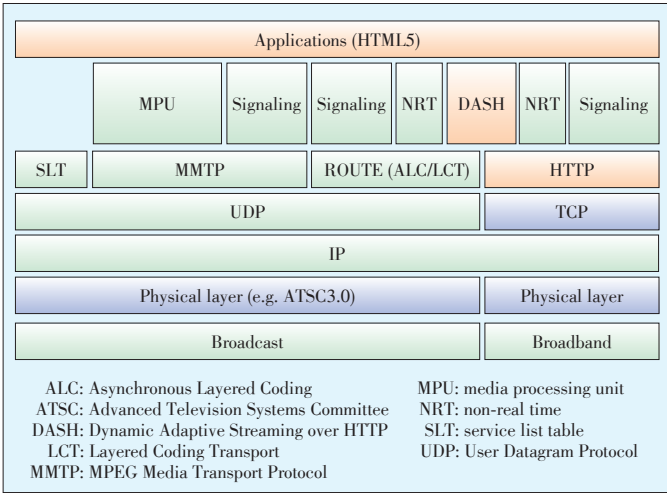
methods, ATSC 3.0 adopts the receiver protocol stack depicted in **Fig. 2**. This structure has the advantages of allowing clean interface among the various layers and enabling the required functionalities via various delivery methods.

In Fig. 2, a typical ATSC 3.0 system contains three functional layers, namely, the physical layer, delivery layer and Application layer. More specifically, the physical layer comprises the broadcast and broadband physical layers. The delivery layer realizes data streaming and object flow transport functionality. The application layer enables various type of services, such as the digital television or HTML5 [17] applications.

ATSC 3.0 incorporates MMT and ROUTE in its Delivery Layer, both of which are carried out via a UDP/IP multicast over the broadcast physical layer. In particular, MMT utilizes the MMT Protocol (MMTP) [13] to deliver media processing units (MPUs), while ROUTE is based on MPEG DASH and Layered Coding Transport (LCT) [18] to deliver DASH segments. The delivery layer of ATSC 3.0 is also equipped with the HTTP protocol [19] with a TCP/IP unicast over the Broadband Physical layer. Non-timed content including the NRT media, electronic program guide (EPG) data, and other files can be delivered with ROUTE or directly over UDP. Signaling may be delivered over MMTP and/or ROUTE, while bootstrap signaling information is the means of the service list table (SLT). To support hybrid service delivery, in which one or more program elements are delivered via the broadband path, MPEG DASH over HTTP/TCP/IP is used on the broadband physical layer. Media files in ISO BMFF are used as the delivery, media encapsulation and synchronization format for both broadcast and broadband deliveries.

**2.2 Functionality of ATSC 3.0 System**

With ATSC 3.0 being a unified IP centric delivery system, content providers have the flexibility of exploring broadcast delivery, broadband delivery, or both to enhance efficiency and potential revenue. For instance, for broadcast services, the us-



▲ **Figure 2. ATSC 3.0 receiver protocol stack.**

## DASH and MMT and Their Applications in ATSC 3.0

Yiling Xu, Shaowei Xie, Hao Chen, Le Yang, and Jun Sun

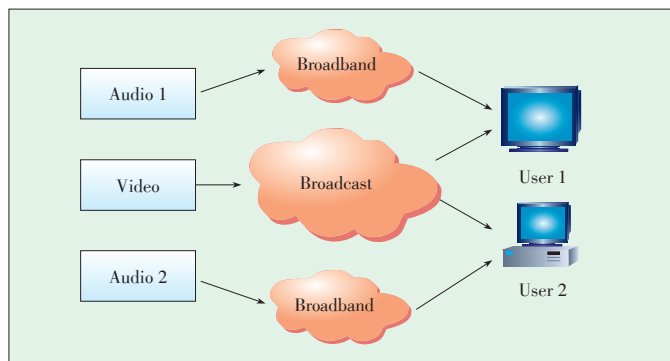
er experience can be improved via a browser-based user interface that enables interactive applications to run in the same playback environment as the streaming media. There are many opportunities for increased revenue from the placement of personalized advertisements. For broadband services, key benefits may be the opportunity to provide narrowly focused content more efficiently than broadcasting. This in turn enables more precise user targeting. There are also opportunities to improve channel capacity via e.g., always utilizing the more efficient delivery method. Therefore, this makes ATSC 3.0 an ideal option for hybrid services [20], which focus largely on either high-penetration contents that are best served by broadcast delivery or contents with narrower interest to be best carried over broadband delivery. ATSC 3.0 can aggregate the content of hybrid services from a variety of sources and deliver it through dynamically exploiting broadcast and broadband distribution channels (Fig. 3).

In addition, with MMT and DASH/ROUTE, services within ATSC 3.0 can be distributed as scalable streams consisting of a base layer and a number of enhancement layers [21]. The broadcaster has different means to transmit these multilayer content. For examples, the base layer content may be delivered via broadcasting, while the enhancement information is transmitted over the broadband network. Alternatively, delivering all layers content only via the broadcast network is also possible.

### 3 Signaling for ROUTE/DASH and MMTP/MPU Service Delivery

#### 3.1 Service List Table and Service Layer Signaling

Signaling information is carried in the payload of IP packets [22] with a well-known address/port and it is commonly referred to as Low Level Signaling (LLS). In ATSC 3.0, service list Table (SLT) is specified as the LLS and its functionality is similar to that of the program association table (PAT) of MPEG-2. To be more specific, SLT supports a rapid channel scan that enables a receiver first encountering the broadcast emission to build a list of all the received ATSC 3.0 services. SLT also pro-



▲ Figure 3. Hybrid services with different delivery channels.

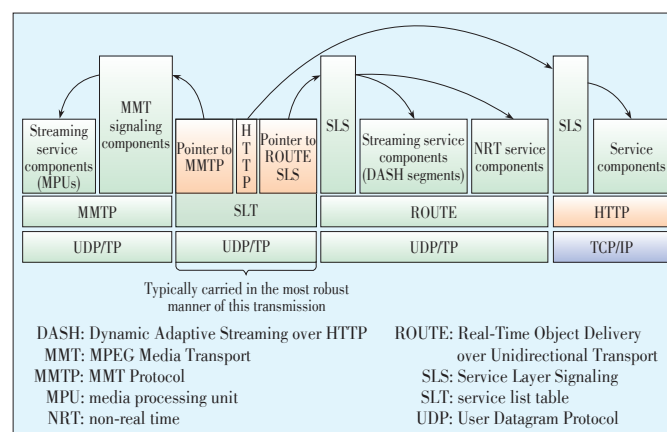
vides bootstrap information that allows a receiver to locate the Service Layer Signaling (SLS). The SLS of each ATSC 3.0 service describes the service characteristics including its content components, where to acquire them as well as the device capabilities needed to produce a meaningful presentation of the service. The bootstrap information provided by the SLT contains the destination IP address and the destination port of the Layered Coding Transport (LCT) session or the MMTP session that carries the SLS for services delivered via ROUTE/DASH or MMTP/MPU.

Fig. 4 summarizes the relationship among the SLT, SLS and the ATSC 3.0 services. It can be seen that for broadcast delivery of ROUTE/DASH services (streaming and NRT), the SLS is carried by ROUTE/UDP/IP in one of the LCT transport sessions, while for the MMTP/MPU streaming services, the SLS is carried by MMTP Signaling Messages. SLS is delivered at a suitable carousel rate to support fast channel join and switching. Under broadband delivery, the SLS is transmitted over HTTP(S)/TCP/IP.

#### 3.2 Service Identification in ROUTE and MMTP Sessions

Each ROUTE session has one or more LCT sessions which carry, as a whole or in part, the content components that make up the ATSC 3.0 service. In streaming service delivery, an LCT session may carry an individual component of a user service, such as an audio, video or closed caption stream. Streaming media is partitioned per MPEG DASH into DASH segments. Each MMTP session contains one or more MMTP packet flows which carry MMT signaling messages for the content components. An MMTP packet flow may carry MMT signaling messages or components formatted per MMT as MPUs.

A ROUTE session is identified via the source IP address, destination IP address and destination port number. An LCT session associated with the service component(s) it carries is identified via the transport session identifier (TSI), which is unique within the scope of the parent ROUTE session. Properties common to the LCT sessions and properties unique to individual LCT sessions are given in a ROUTE signaling structure



▲ Figure 4. SLT references to services via SLS.

## DASH and MMT and Their Applications in ATSC 3.0

Yiling Xu, Shaowei Xie, Hao Chen, Le Yang, and Jun Sun

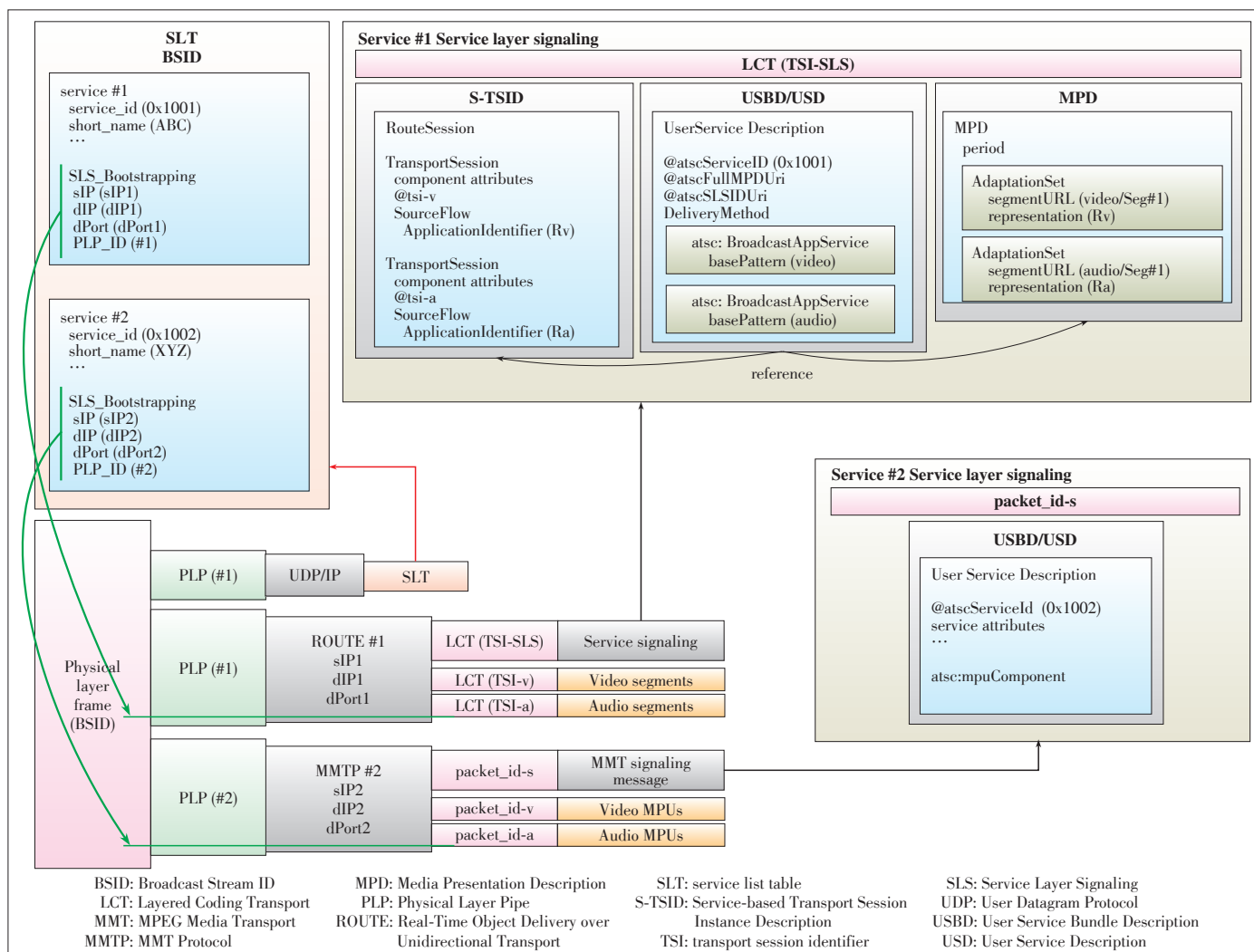
called Service-based Transport Session Instance Description (S-TSID), which is part of the SLS. Each LCT session is carried over a single physical layer pipe. Different LCT sessions of a ROUTE session may or may not be contained in different physical layer pipes. The properties described in the S-TSID include TSI value and PLP\_ID for each LCT session, descriptors for the delivery objects/files, and Application Layer Forward Error Correction (AL-FEC) parameters [23].

The MMTP session and packet flow are identified in a similar manner as the ROUTE session. In particular, the destination IP address/port number and the packet\_id unique within the scope of the parent MMTP session are used. Properties common to each MMTP packet flow and properties of individual MMTP packet flows are included in the SLT. Properties for each MMTP session are given by MMT signaling messages, which may be carried within the MMTP session. These properties include packet\_id value and PLP\_ID for each MMTP packet flow.

Fig. 5 gives an example on the use of SLT to bootstrap the

SLS acquisition and also the use of the SLS to acquire service components delivered on either ROUTE sessions or MMTP sessions. In this example, ATSC 3.0 receiver starts acquiring the SLT and each service identified by service\_id provides the following SLS bootstrapping information: PLP\_ID, source IP address (sIP), destination IP address (dIP) and destination port number (dPort).

For service delivery in ROUTE, the SLS consists of the following metadata fragments: User Service Bundle Description (USBD), S-TSID and DASH Media Presentation Description (MPD). These are all pertinent to a certain service. The USBD/USD fragment contains service identification, device capability information, Uniform Resource Identifier (URI) [24] references to the S-TSID fragment and the MPD fragment, and metadata to enable the receiver to determine whether the transport mode is broadcast and/or broadband for service components. The S-TSID fragment provides component acquisition information associated with a service and mapping between DASH representations found in the MPD and the TSI corresponding to the



▲ Figure 5. Exemplary use of service signaling for bootstrapping and service discovery.

component of the service.

For service delivery using MMTP, the SLS contains only USBD fragments. These mainly reference the MMT signaling's MMT package table (MPT) message that provides identification of package ID and location information for assets belonging to the service. MMT signaling messages are delivered by MMTP according to the signaling message mode specified in [13].

To attain compatibility with MPEG DASH components, MMTP SLS needs to include USBD fragments used in ROUTE/DASH and even MPD fragments. This would lead to signaling redundancy. It is still unclear whether MPD should be carried as an SLS metadata fragment in ROUTE/DASH or in the payload of media presentation information (MPI) message, which is an instance type of MMT signaling messages. On the other hand, due to the USBD fragment being reused, USBD for MMT can give complete service bundle description for both media component formats, i.e., the DASH segment and MPU format. This means that MMTP/MPU can support the delivery of both media formats. As a result, MMTP/MPU is more inclusive and robust to diverse media component formats. In addition, MMTP/MPU defines more detailed USBD information about the service, including attributes such as @atsc:providerId, @atsc:ServiceCategory and @atsc:serviceStatus, and elements such as atsc:channel and atsc:componentInfo. At the cost of increased redundancy, MMTP/MPU signaling can achieve richer and more comprehensive media service.

## 4 Content Delivery

### 4.1 Overview

MPEG DASH is designed for communication between HTTP servers and DASH client(s). It specifies formats and methods for delivering streaming service(s), and describing a collection of media segments and auxiliary metadata through an MPD. The ROUTE protocol provides general broadcast delivery capabilities similar to the broadband delivery capabilities provided by HTTP. In addition to supporting the file-based streaming techniques used in DASH, ROUTE provides media-aware content delivery, which enables faster channel switch. On the whole, ROUTE/DASH is a broadcast-oriented, media-aware byte range delivery protocol based on MPEG DASH and LCT over UDP, with the use of AL-FEC. ROUTE/DASH can also operate with any media codec, including scalable media codecs provided that the appropriate codec specific file format is specified.

MMTP is a protocol designed to deliver ISOBMFF files and it is very useful in real-time streaming delivery via an unidirectional delivery network. MMTP has several distinct features such as:

- media-aware packetization of ISOBMFF files
- multiplexing of various media components into a single MMTP

session

- network jitter removal at the receiver under the constraints set by the sender
- receiver buffer management by the server to avoid any buffer underflow or overflow and the fragmentation/aggregation of payload data
- detection of missing packets during delivery.

It is noteworthy that in ATSC 3.0, the use of AL-FEC in ROUTE is crucial for e.g. the large NRT file delivery, DVR applications and enhanced robustness for collecting small objects, while in MMTP, it is optional.

### 4.2 Media Encapsulation

In broadcast service, since ROUTE/LCT doesn't differentiate DASH Segments by type, it may introduce more service delay considering the dependency among the segments. On the contrary, MPU is self-contained, i.e., the initialization information and metadata required to fully decode the media data in each MPU is included in the MPU. In MMTP sessions, the media fragment type of the payload is known, leading to easy construction of the media data. These two properties of MMTP/MPU improve system performance in terms of resistance and robustness. For instance, with the knowledge on the fragment type in MMTP session, adaptive AL-FEC protection could be used on the basis of the significance of media fragments to achieve better user experience. In terms of random access and channel switch, MMTP/MPU acquires the service in MPU level faster due to self-contained attributes. By comparison, ROUTE/DASH may force users to wait until the next metadata segment arrives.

In addition, with the concept of Media Delivery Event (MDE), ROUTE/DASH users need more time to restore a service when abrupt data loss or error occurs in MDE starting with a Random Access Point (RAP), until the next MDE starts with a RAP. Another important issue is that live advertisement insertion and removal are easier for MMTP/MPU because there is no structural difference between an ad MPU and a program MPU.

Finally, for MMTP/MPU, each MPU contains a globally unique ID for media components and a sequence number to enable unique identification of each MPU regardless of the delivery mechanism. As for DASH, media segments and auxiliary metadata are referenced by Uniform Resource Locator (URL) through MPD, while in a ROUTE session, S-TSID provides the mapping between DASH representations found in the MPD and the TSI corresponding to the component of the service. The identification scheme used in MPU is more accurate and scalable, which benefits media allocation and recognition in media service.

### 4.3 Streaming Services

Flexible packaging and diverse delivery modes supported both in ROUTE/DASH and MMTP/MPU enable multiple types

### DASH and MMT and Their Applications in ATSC 3.0

Yiling Xu, Shaowei Xie, Hao Chen, Le Yang, and Jun Sun

of service components to be packaged and delivered in a way best for them. This feature enhances agility and effectiveness of delivery method and it also dramatically improves the QoS for the system and QoE of users.

#### 4.3.1 Delivery in ROUTE/DASH

In the streaming service of ROUTE/DASH, no matter it is live content or pre-recorded content, the attribute ‘type’ of the MPD (MPD@type) should be set to “dynamic”. As for the attribute ‘minimumUpdatePeriod’ (MPD@ minimumUpdatePeriod), when it is present, the receiver should get MPD updates carried in the LCT session. The objects delivered by LCT session of the ROUTE protocol shall be formatted according to the announcement in the MPD. The MPD and the described Media Presentation should conform to the ISO Base media file format as specified in [8].

In streaming services delivery using ROUTE/DASH, three different kinds of delivery mode are proposed, namely, Entity Mode, File Mode and Packaging Mode. The Entity Mode is used when it is not possible to determine the Extended File Delivery Table (EFDT) parameters in ROUTE prior to the object delivery. In this case, the EFDT parameters (as entity-headers) are sent in-band with the delivered object (as the entity-body) in the form of a compound object. The file/object metadata is carried by one or more entity-header fields associated with the entity-body. Meanwhile, Entity Mode enables partial or chunked delivery in the same way as HTTP, and it provides the means to reduce the sender delay and possibly the end-to-end delay as well. In the File Mode, the file/object metadata as represented by the EFDT would either be embedded within or be referenced as a separate delivery object. The file may also be sent in a progressive manner using the regular ROUTE sending operation in order to reduce sender delay. The Packaging Mode should be used as the delivery object format if the repair flow is used in conjunction with the source flow for streaming content delivery. It enables more robust AL-FEC recovery by applying FEC protection across a collection of delivered objects for enhanced time diversity and constant QoS.

#### 4.3.2 Delivery in MMTP/MPU

Each content component is considered an MMT asset under MMTP/MPU. Each MMT asset is a collection of one or more MPUs with the same unique Asset ID. An MMT package is a collection of one or more assets, and an ATSC 3.0 Service can have one or more MMT packages. Both MMT packages and MPUs do not overlap in their presentation time.

Multiple assets can be delivered over a single MMTP session. Each asset is associated with a packet\_id which is unique within the scope of the MMTP session. This enables efficient filtering of MMTP packets carrying a specific asset. Additionally, MMT signaling messages delivered to the receiver are used to designate the mapping information between MMT packages and MMT sessions.

Fig. 6 shows exemplary mapping between a MMT package and an MMTP session. The MMT package has three assets: asset A, asset B and asset C and is delivered over two MMTP sessions. Although all the MMT signaling messages required to consume and deliver the MMT package should be delivered to the receiver, only a single MPT message is shown for simplicity. Alternatively, all the MMT assets of the MMT package and its associated MPT message can be multiplexed into a single MMTP session, as in the configuration of MPEG-2 TS, which together with the inclusion of the packet\_id field valid range and Clock Relation Information (CRI) message makes MMTP/MPU backward-compatible with MPEG-2 TS.

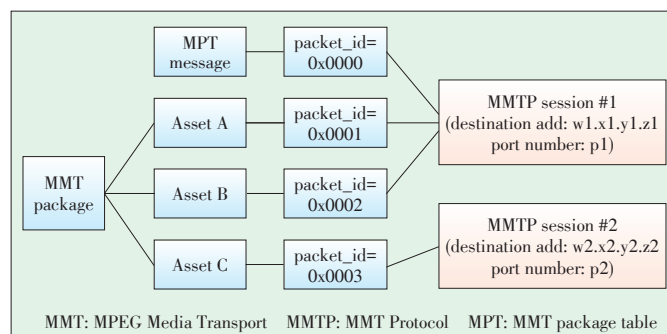
In addition to a single MMT Package being able to be delivered over one or more MMTP sessions, multiple Packages may be delivered by a single MMTP session and multiple packages in the same service can be delivered simultaneously. MMTP provides the media-aware packetization of ISOBMFF files for real-time streaming delivery via an unidirectional delivery network.

#### 4.4 NRT Content

In ATSC 3.0, the media content includes streaming service and NRT content. Streaming service can be delivered by either ROUTE/DASH or MMTP/MMT whereas NRT content can only be delivered over ROUTE/DASH because Generic File Delivery (GFD) mode specified in [6] cannot be used in ATSC 3.0 system.

File content comprising discrete media are considered to be NRT content. When delivering this kind of file content, the File Mode of ROUTE/DASH should be used. Before the delivery of the file, the file metadata (EFDT) should be informed to the receiver. In the delivery of the subsequent source flow, the delivered file references the LCT session, and the file Uniform Resource Identifier (URI). The file URI is used to identify the delivered EFDT. In this case, the receiver can use the EFDT to process the content file.

Besides common NRT content, service metadata belonging to an NRT content item of an ATSC 3.0 service can also be considered as NRT content (e.g., the SLS or Electronic Service Guide (ESG) fragments) from the application transport and IP delivery perspective. Their delivery also conforms to the princi-



▲ Figure 6. A package delivered over two MMTP sessions.



ples described above.

#### 4.5 Synchronization

Regarding synchronization, DASH takes advantage of UTC [25] to fulfill the accurate requirement of wall clock. Since UTC can be established over the physical layer, a receiving device can obtain wall clock when connecting to ATSC 3.0 broadcast. In addition, servers of the same service can synchronize with respect to a common wall clock (UTC) source via broadcast or broadband networks. Notably, the broadcast - established wall clock is mainly used by servers for serving media components at the receiver.

In MMT, the synchronization of MPUs is also based on timestamps referencing UTC. The MPU\_timestamp\_descriptor as defined in [13] is used to represent the presentation time of the first media sample in terms of the presentation order in each MPU.

#### 4.6 System Model for Media Delivery

##### 4.6.1 ROUTE/DASH System

The generic ROUTE/DASH system model is shown in **Fig. 7**. ROUTE/DASH relies on the concept of discrete - time events, e.g.,  $m$  bytes input to or  $n$  bytes output from the buffers at a specific instant. This results in no leakage rates specified. There is a transport buffer (TB<sub>n</sub>) for a specific ROUTE session that may deliver multiple objects and related AL-FEC as encapsulated by ROUTE/UDP/IP. Objects for media services to be delivered are briefly stored in the ROUTE output buffer before they are consumed by the DASH client. The minimum size of this buffer is slightly smaller than that of the associated TB<sub>n</sub>, as the data in TB<sub>n</sub> would be wrapped in ROUTE/UDP/IP and may include AL-FEC packets. The output objects/files are decapsulated and decoded.

The EBN buffer is defined in MPEG systems as an elementary stream buffer. This buffer, when applied to ROUTE/DASH,

is integrated with the ISO BMFF file handler that holds data until it is parsed to the decoders. Given that there may exist multiple object/file streams in an LCT session, there may be several EBN(s) associated with a given LCT session. There may also be multiple media types within a single ISOBMFF file stream. As such, there are multiple decoders D<sub>n</sub> connected to each EBN.

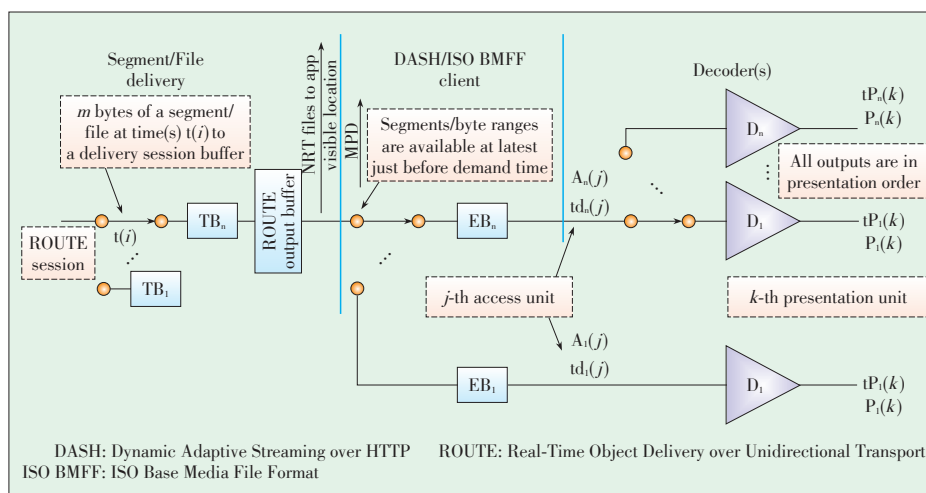
The task of scheduling media to the codec(s) at the receiver is handled solely by the ISOBMFF file handler (within the DASH client). Scheduling objects/files to the ISOBMFF handler is part of the DASH function. It consists of a series of steps that lead to the ISOBMFF handler receiving media delivery events (MDEs) that include byte ranges or object(s) that make contextual and temporal sense, and allow the handler to deliver samples (media frames) to the codecs to fulfill the media presentation timeline.

The operation of the ROUTE/DASH system is defined such that none of the constituent buffers, namely, TB<sub>n</sub>, Ebn and the ROUTE Output Buffer, are allowed to overflow, and codec(s) cannot stall due to lack of input media data. Each buffer has no data in the initialization stage and may likely become empty briefly during system operation. A notable aspect of the ROUTE/DASH system model is the lack of physical layer buffer. This is crucial for enabling ROUTE/DASH to perform at or close to the theoretical limit of the channel variation rate. ROUTE Transport Buffer Model is a loop-locked subsystem in which each module can proactively send feedback to others if needed without external assistance. This also means no extra signaling messages are necessary to support the operation of ROUTE transport buffer subsystem.

##### 4.6.2 MMTP/MPU System

**Fig. 8** shows the procedure of the content delivery, acquisition and playback of service in MMTP/MPU, which can be describe in the following steps:

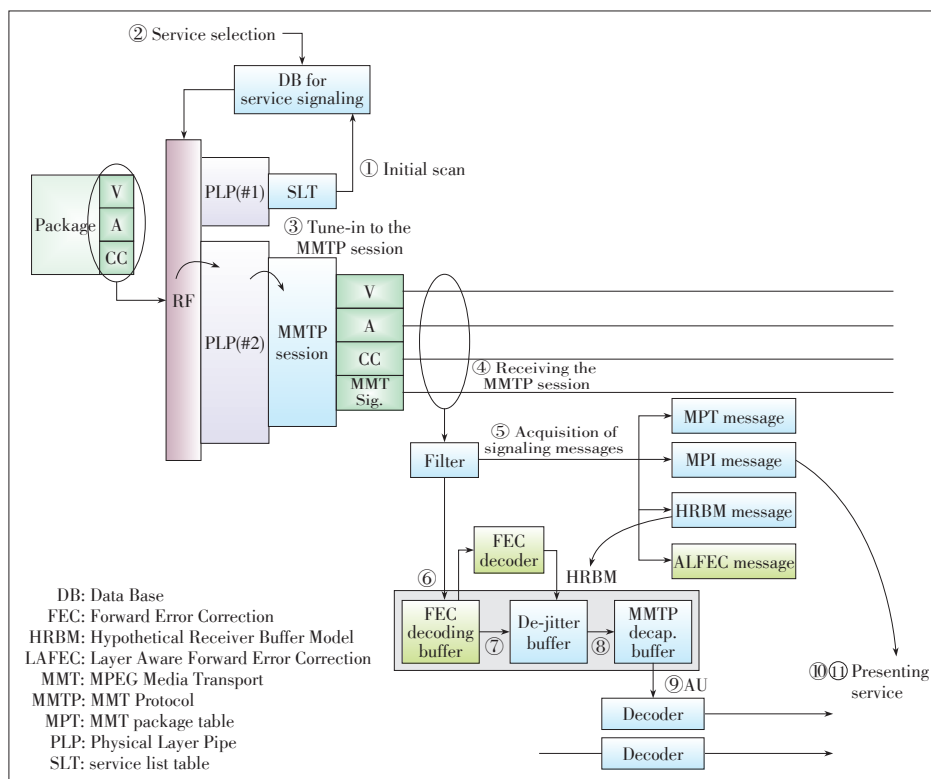
- 1) The client acquires a service list table (SLT) to scan the channel information and service signaling messages to attain the service level information.
- 2) The user selects a service and identifies the corresponding MMTP\_package\_id.
- 3) The channel information acquired in step 1) is used by an RF channel, and the PLP selection to precede the service selection. According to the service selection, an MMTP session carrying the corresponding MPT message is acquired.
- 4) In the referenced MMTP session, various content components and signaling messages are obtained through a few MMTP packets.



▲ **Figure 7.** System model for ROUTE/DASH delivery.

## DASH and MMT and Their Applications in ATSC 3.0

Yiling Xu, Shaowei Xie, Hao Chen, Le Yang, and Jun Sun



▲ Figure 8. System model for MMTP/MPU delivery.

- 5) The type field of MMTP packet headers can inform the receiver to obtain corresponding signaling messages. The MPT table extracted from the MPT message is processed to get the list of MMT assets comprising the selected service with the designated MMT\_package\_id. Other MMT signaling messages are processed if necessary, including the MPI, Hypothetical Receiver Buffer Model (HRBM) and AL-FEC.
- 6) From the MMTP packets received, different MMTP packets corresponding with each content component are filtered and stored in their corresponding FEC decoding buffers. MMTP packets with repair symbols corresponding to each content component are also received and stored separately.
- 7) MMTP packets received at the FEC Decoding Buffer are immediately copied into a corresponding de-jitter buffer. Missing packets can be detected using the packet\_sequence\_number of each MMTP packet. If a packet is not received before a predefined time specified by an AL-FEC message, it is considered missing and should be recovered by applying the AL-FEC code. The recovered packet needs to be copied to the de-jitter buffer immediately.
- 8) HRBM field specifies the amount time MMTP packets should spend in the de-jitter buffer.
- 9) MMTP packets of MPUs are processed to extract Access Units.
- 10) The first AU in an MPU is decoded by the appropriate decoder and presented at the time designated by the MPU\_timestamp.

11) The next AU in the MPU is decoded and presented after the presentation of the first AU. This step is repeated until the last AU of the MPU has been decoded and presented.

For streaming service using MMTP/MPU, each MMTP packet with media data has an explicit indication of the boundaries of the media samples or sub-samples. As a result, MMTP packets with media data only carry minimum information about media data (such as a movie fragment sequence number and a sample number) needed to recover the association between the media data and the metadata. By contrast, ROUTE/DASH considers media segments just as payload, which indicates that more supplementary design in signaling and buffer model are needed.

In an MMTP/MPU system, HRBM is introduced so that a broadcast server can emulate the behavior of the receiver buffer, and any processing the receiver performs on the packet streams is within the reception constraints of the receiver.

The HRBM message is signaled from the server to a client to guide the operation of receiver buffer subsystem. HRBM may shift more workload to the server-side, and it is possible that the subsystem does not work well in practice (e.g., buffer overflow or underflow may occur since the information provided in HRBM message may not match diverse environment situations). Moreover, the goal of HRBM is to achieve constant delay for the delivery system of the MMT receiving entity to synchronize the presentation of the media data, but this guideline may not apply for the situation that there exist vast delay gap among several transmission paths or networks. The de-jitter buffer of HRBM can mitigate the jitter introduced by multipath, multisource or multi-network though.

### 4.7 Rules for Session Presence

In ATSC 3.0, the rules regarding the presence of ROUTE/LCT sessions and/or MMTP sessions for carrying the content components of an ATSC 3.0 service are as follows:

- 1) For broadcast delivery of a linear service without application-based enhancement, the service's content components are carried by either
  - One or more ROUTE/LCT sessions, or
  - One or more MMTP sessions.
- 2) For broadcast delivery of a linear service with application-based enhancement, the service's content components are carried by:
  - One or more ROUTE/LCT sessions, and

- Zero or more MMTP sessions.

The use of both MMTP and ROUTE for streaming media components in the same service is not allowed, and the selected protocol is specified in SLT.

- 3) For broadcast delivery of an application-based service, the content components are carried by:

- One or more ROUTE/LCT sessions.

The combination of sessions with the NRT content can only be carried by one or more ROUTE/LCT sessions (as described above). ROUTE/LCT can deliver almost all types of service components without the assistance of MMTP, but the inverse is not possible. Using ROUTE/LCT for delivery of service components appears to be more systematic than MMTP, while the latter is an alternative delivery solution for linear service in ATSC 3.0.

## 5 Hybrid Delivery Mode

Hybrid broadcast broadband TV (HbbTV) [26] is a globally initiative technology mainly developed by industry leaders. HbbTV aims at improving user experience when it comes to hybrid content, by harmonizing the broadcast and broadband delivery through connected TVs, set-top boxes and multiscreen devices. Combining elements of existing standards, including OIPF, CEA, DVB, MPEG-DASH and W3C, HbbTV specification defines a hybrid (broadcast and broadband) platform for the signaling, transport and presentation of enhanced or interactive applications on hybrid terminals, which include both a DVB compliant broadcast connection and a broadband connection to the Internet. In HbbTV, legacy DVB broadcast standards and systems are furthest maintained, while parts of already available standards based on IP network are largely referenced and adapted where necessary.

In order to integrate all the above technologies organically, HbbTV specification may need to afford intricate schemes due to the compatibility of data format, processing procedure and system architecture. Relatively, the design and development in ATSC 3.0 system are all based on IP network, and the transport protocols in application and physical layers are redesigned considering new requirements and scenarios. Therefore, ATSC 3.0 system could provide more systematic and fundamental solutions to support hybrid delivery of enhanced or interactive services.

This section describes hybrid delivery modes. One possible hybrid delivery operation employs ROUTE/DASH in the broadcast path and DASH over HTTP(s) in the broadband path. The other mode uses MMTP/MPU in the broadcast path and DASH over HTTP(s) in the broadband path.

### 5.1 Mode with ROUTE/DASH in Broadcast

The hybrid service delivery mode with ROUTE/DASH combines the broadcast delivery via ROUTE, and unicast delivery via HTTP. In some scenarios, e.g., due to device movement,

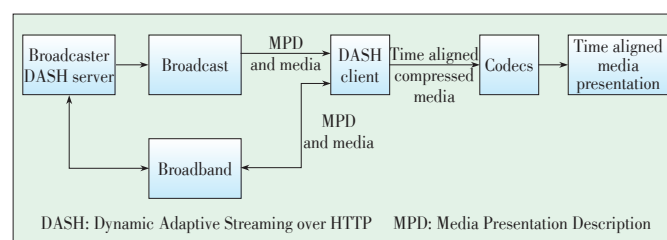
the broadcast signal may become unavailable, which would result in handover to broadband and may involve a subsequent return to the broadcast service. **Fig. 9** shows the hybrid mode operation within the DASH client.

The presence of an unmanaged broadband network is likely to introduce a variable amount of extra delay (latency) in the delivery, as the network congestion can impact the availability of the broadband-delivered content. ROUTE/DASH allows the broadcaster to employ techniques specified in the DASH standard to improve user experience in the hybrid delivery case. Buffering is very important and is handled in a slightly different way for different delivery modes. DASH segments for unicast broadband components are requested by receivers ahead of when they are needed, as indicated by the timeline set by the DASH MPD, and they are buffered until their presentation time. The broadcaster can ensure that broadband-delivered service components are made available to the broadcaster server well ahead of the broadcast-delivered components of the service to accommodate slower connections.

The receivers control delivery timing and buffering. Segments for broadcast components are delivered from the broadcast source according to the buffer model, which ensures that receivers get the segments soon enough to avoid decoder stall but not so soon as to reduce buffer overflow. The broadcast source uses the DASH MPD timeline to determine the appropriate delivery timing.

The possible scenario of switching from broadcast to broadband service access or vice versa involves the mobile ATSC 3.0 receiver, which due to user mobility, may move temporarily outside the broadcast coverage and only fallback service reception via broadband is possible. Support for such handover between different access modes may be indicated by the MPD fragment in SLS. In this case, as defined by the SLS protocols, the `userServiceDescription.deliveryMethod` element in the USBD fragment contains the child elements `atsc:broadcastAppService` and `atsc:unicastAppService`, which represent the broadcast and broadband delivered components of the named service. These broadcast and unicast delivered components may be substituted for one another in the case of handover from broadcast to broadband service access and vice versa.

Thanks to the uniformity of data encapsulation format in ROUTE/DASH, the system design and realization adapted for streaming service with app-based enhancement is simple and systematic, which means that there is few conversion in the me-



▲ **Figure 9.** ROUTE/DASH hybrid delivery mode.

## DASH and MMT and Their Applications in ATSC 3.0

Yiling Xu, Shaowei Xie, Hao Chen, Le Yang, and Jun Sun

dia and signaling formats.

### 5.2 Mode with MMTP/MPU in Broadcast

The hybrid streaming in this mode over broadcast and broadband delivery is shown in **Fig. 10**. All the components of the system are locked to UTC for synchronization.

For the broadcast network, media data are encapsulated into MPUs, which are packetized into MMTP packets. For the broadband network, media data are encapsulated into DASH segments. When no components are delivered by broadcast, DASH MPDs are delivered over the broadcast network by a signaling message, and DASH segments are delivered via broadband by an HTTP session through the network interface of a regular HTTP server. This solution increases the system complexity and operational cost.

For the client, it is assumed that the MMTP packets delivered through the broadcast network are de-packetized and the media data are decoded by the appropriate media decoders, and that the DASH segments are delivered through the broadband network. To synchronize the presentation of a DASH segment delivered via the broadband network with an MPU delivered via the broadcast network, the presentation time of the DASH segment is represented by a timestamp referencing UTC.

Like the mode with ROUTE/DASH used in broadcast, when it comes to the switch from broadcast to broadband, the client uses, before the transition, the latest MPD contained in an MPI message to make an HTTP request for a DASH segment. The received DASH segment is buffered and made available to the DASH client for decoding. For seamless handoff, the broadcast delivery delay should be adjusted so that it is equal to the broadband delivery delay; otherwise, service may freeze when switching from broadcast to broadband for the first time. Meanwhile, the DASH segment and MPU are inconsistent with each other in format, resulting in more processing time and delay in data conversion. Also, for handoff services, MPU boundary may overlap with DASH segment boundary, which can create barriers to seamless handoff and further degrade the service quality.

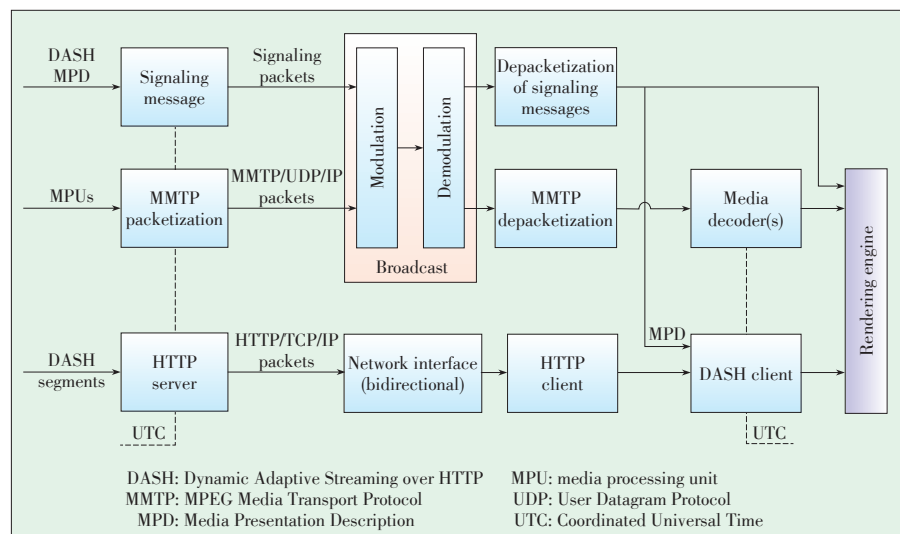
Because a transition can only happen on an MPU boundary, at least one DASH segment whose boundary matches with broadcast MPUs needs to be buffered in advance. In **Fig. 11**, the client has detected a loss in broadcast signaling for MPU  $n$ , and makes an HTTP request for the DASH segment cor-

responding to the lost  $n$ th MPU. There is no transition delay because MPU  $n-1$  is presented during the time the client is receiving the designated DASH segment. The HRBM is used by the broadcaster to add a specific amount of delay required for the seamless transition. When the broadcast signal becomes available again, there is a seamless transition to the broadcast in this example.

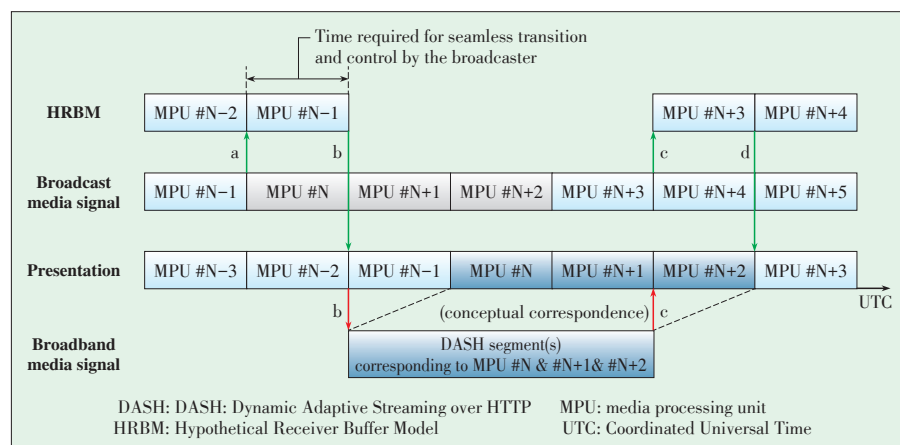
## 6 Conclusions

This paper described the features and design considerations of ATSC 3.0, a next-generation broadcasting system designed to address the emerging multimedia service delivery requirements with using broadcasting channels in combination with broadband networks. It also provided comparative comparative introductions and applications in details about ROUTE/DASH and MMTP/MPU adopted in the transport protocols used in ATSC 3.0 for broadcasting.

In ATSC 3.0, the ROUTE/DASH is derived from FLUTE



▲ Figure 10. MMTP/MPU hybrid delivery mode.



▲ Figure 11. Seamless transition: broadcast-broadband-broadcast.



## DASH and MMT and Their Applications in ATSC 3.0

Yiling Xu, Shaowei Xie, Hao Chen, Le Yang, and Jun Sun

[27] and DASH, FLUTE is designed for NRT data push services over unidirectional transport and DASH was developed to realize dynamic adaptive streaming over HTTP. It is the first time that ROUTE/DASH has been applied to a broadcasting system. MMTP/MPU was accepted as the MMT standard in 2012. In Japan, super hi-vision test services are scheduled to begin in 2016, and commercial services are scheduled to begin in 2020 using MMT as a transport protocol for next-generation broadcasting systems.

As developments of ROUTE/DASH and MMT standards, as well as next generation broadcasting system is still in progress, further study and verification of technologies adopted in ATSC 3.0 are still needed to be done.

## References

- [1] *2nd MMT Workshop in Kyoto: Presentations*, ISO/IEC JTC 1/SC 29/WG 11 N11200, MPEG, 2010.
- [2] P. Podhradsky, "Evolution trends in Hybrid Broadcast Broadband TV", *55th International Symposium ELMAR-2013*, Zadar, Yugoslavia, Sept. 2013.
- [3] *Information Technology — Generic Coding of Moving Pictures and Associated Audio Information: Part 1 Systems*, ISO/IEC 13818-1, 2013.
- [4] *Information Technology — Coding of Audio-Visual Objects — Part 12 ISO Base Media File Format*, ISO/IEC 14496-12, 2012.
- [5] *Digital Video Broadcasting (DVB), Generic Stream Encapsulation (GSE) Protocol*, ETSI TS 102 606 V1.1.1, Oct. 2007.
- [6] R. Pantos and E. W. May. (2011, Oct. 2). HTTP Live Streaming [Online]. Available: <http://tools.ietf.org/html/draft-pantos-http-live-streaming-06>
- [7] Microsoft. (2009, Sep. 8). IIS Smooth Streaming Transport Protocol [Online]. Available: <http://www.iis.net/learn/media/smooth-streaming/smooth-streaming-transport-protocol>
- [8] *Information technology — Dynamic adaptive streaming over HTTP (DASH) — Part 1: Media presentation description and segment formats*, ISO/IEC 23009-1, May 2014.
- [9] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the internet," *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, Apr. 2011. doi: 10.1109/MMUL.2011.71.
- [10] *ATSC 3.0 Management and Protocols – Signaling, Delivery, Synchronization, and Error Protection*. ATSC Working Draft S33-174r0, Dec. 2015.
- [11] T. Stoekhammer, "Dynamic adaptive streaming over HTTP-design principles and standards," in *Proc. Second Annual ACM Conference on Multimedia Systems*, New York, USA, 2011, pp. 2–4.
- [12] C. Müller, S. Lederer, and C. Timmerer, "An evaluation of dynamic adaptive streaming over HTTP in vehicular environments," in *Proc. 4th Workshop on Mobile Video*, New York, USA, 2012, pp. 37–42. doi: 10.1145/2151677.2151686.
- [13] *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 1: MPEG media transport (MMT)*, ISO/IEC 23008-1, 2nd Edition, Jun. 2015.
- [14] S. Aoki, K. Otsuki, and H. Hamada, "New media transport technologies in super hi-vision broadcasting systems," in *Proc. International Broadcasting Convention*, Amsterdam, Netherlands, Sept. 2013, pp. 7–9. doi: 10.1049/ibc.2013.0029.
- [15] Y. Lim, K. Park, J. Y. Lee, S. Aoki, and G. Fernando, "MMT: an emerging MPEG standard for multimedia delivery over the internet," *IEEE Multimedia*, vol. 20, no. 1, pp. 80–85, Mar. 2013. doi: 10.1109/MMUL.2013.7.
- [16] S. Aoki, K. Otsuki, and H. Hamada, "Effective usage of MMT in broadcasting systems," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, London, UK, Jun. 2013, pp. 1–6. doi: 10.1109/BMSB.2013.6621756.
- [17] W3C. HTML5: A vocabulary and associated APIs for HTML and XHTML [Online]. Available: <http://www.w3.org/TR/2014/REChTML5-20141028/>
- [18] *Layered Coding Transport (LCT) Building Block*, IETF: RFC 5651, Oct. 2009.
- [19] *Hypertext Transfer Protocol – HTTP/1.1*, IETF: RFC 2616, Jun. 1999.
- [20] Y. Lim, S. Aoki, I. Bouazizi, and J. Song, "New MPEG transport standard for next generation hybrid broadcasting system with IP," *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 160–169, Jun. 2014. doi: 10.1109/TBC.2014.2315472.
- [21] C. Müller, D. Renzi, S. Lederer, et al., "Using scalable video coding for dynamic adaptive streaming over HTTP in mobile environments," in *Proc 20th European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 2208–2212.
- [22] S. Aoki and K. Aoki, "Efficient multiplexing scheme for IP packets over the advanced satellite broadcasting system," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 1, pp. 49–55, Feb. 2009. doi: 10.1109/TCE.2009.4814413.
- [23] *Forward Error Correction (FEC) Framework*, IETF: RFC 6363, Oct. 2011.
- [24] Joint W3C/IETF URI Planning Interest Group. (2001, Sept. 21). URIs, URLs, and URNs: Clarifications and Recommendations 1.0 [Online]. Available: <https://www.w3.org/TR/uri-clarification/>
- [25] W3C. (1997, Sept. 15). Date and Time Formats [Online]. Available: <http://www.w3.org/TR/1998/NOTE-datetime-19980827>
- [26] *Hybrid Broadcast Broadband TV*, ETSI TS 102 796 V1.3.1, Oct. 2015.
- [27] *FLUTE - File Delivery over Unidirectional Transport*, IETF: RFC 6726, Nov. 2012.

Manuscript received: 2015-11-15

## Biographies

**Yiling Xu** (yl.xu@sjtu.edu.cn) received her PhD in electrical engineering from the University of Electronic Science and Technology of China. From 2004 to 2013, she was with Multimedia Communication Research Institute of Samsung Electronics Inc, where she focused on digital broadcasting, IPTV, convergence networks and next-generation multimedia applications. Currently, she is with the Shanghai Jiao-tong University, as a research associate. Dr. Xu has more than 70 patents and 10 academic papers. She is active in international standard organizations including DVB, MPEG, 3GPP, OIPE, OMA, and FOBTV.

**Shaowei Xie** (sw.xie@sjtu.edu.cn) received his BE degree in electronics and information engineering from Northwestern Polytechnical University, China in 2014. He is pursuing his PhD degree at the Institute of Image Communication and Signal Processing, Shanghai Jiao Tong University, China. His research interest is multimedia communication.

**Hao Chen** (chenhao1210@sjtu.edu.cn) received his BE degree in electronics and information engineering from Northwestern Polytechnical University, China in 2013. He is pursuing his PhD degree at the Institute of Image Communication and Signal Processing, Shanghai Jiao Tong University, China. His research interest is multimedia communication.

**Le Yang** (le.yang.le@gmail.com) received his BEng. and MSc degrees in electrical engineering from the University of Electronic Science and Technology of China, China in 2000 and 2003, respectively. He received his PhD degree in electrical and computer engineering from the University of Missouri, USA in 2010. Since 2011, he has been an associate professor with Jiangnan University, China. His research interests include sensor networks, passive localization, tracking and signal detection.

**Jun Sun** (junsun@sjtu.edu.cn) received his BS and MS degrees from the University of Electronic Science and Technology of China in 1989 and 1992, respectively, and the PhD degree from Shanghai Jiao Tong University, China in 1995. He is a professor with the Institute of Image Communication and Information Processing of Shanghai Jiao Tong University, China. His research interests include image communication, HDTV, and mobile communication.



# Introduction to AVS2 Scene Video Coding Techniques

Jiaying Yan<sup>1,2,3</sup>, Siwei Dong<sup>1,3</sup>, Yonghong Tian<sup>1,3</sup>, and Tiejun Huang<sup>1,3</sup>

(1. National Engineering Laboratory for Video Technology, School of EE & CS, Peking University, Beijing 100871, China;

2. School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, Shenzhen 518055, China;

3. Cooperative Medianet Innovation Center, Beijing 100871, China)

## Abstract

The second generation Audio Video Coding Standard (AVS2) is the most recent video coding standard. By introducing several new coding techniques, AVS2 can provide more efficient compression for scene videos such as surveillance videos, conference videos, etc. Due to the limited scenes, scene videos have great redundancy especially in background region. The new scene video coding techniques applied in AVS2 mainly focus on reducing redundancy in order to achieve higher compression. This paper introduces several important AVS2 scene video coding techniques. Experimental results show that with scene video coding tools, AVS2 can save nearly 40% BD-rate (Bjontegaard-Delta bit-rate) on scene videos.

## Keywords

AVS2; scene videos coding; background prediction

## 1 Introduction

The primary application of AVS2 is in ultrahigh-definition videos, especially scene videos. Scene videos are usually captured by stationary cameras and include videos from surveillance systems all over the world and from other applications, such as video conference, online teaching and remote medical. Scene videos have huge temporal and spatial redundancy for the background regions appear frequently and AVS2 can utilize the background information to compress the scene videos efficiently.

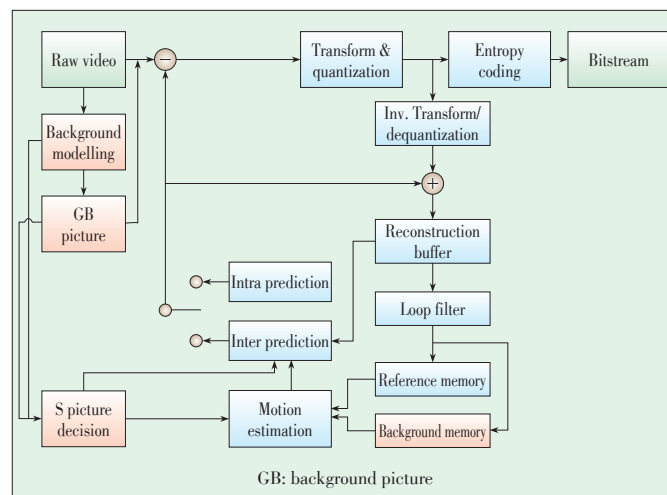
Similar to previous coding standards, AVS2 still adopts the classic block-based hybrid video framework. However, in order to improve coding efficiency, in the AVS2 coding framework, a more flexible coding unit (CU), prediction unit (PU) and transform unit (TU) based structure is adopted to represent and organize the encoded data. With the quad-tree structure, the sizes of CUs are various from  $8 \times 8$  to  $64 \times 64$ . At the same time, the PUs are not limited to symmetric partition while asymmetric PUs are also available. To make coding more flexible, the size of TUs is independent from the size of PUs. Moreover, creative techniques are adopted in AVS2 modules of prediction, transform, entropy coding, etc. [1]. **Fig. 1** describes the video coding architecture.

This work is partially supported by the National Basic Research Program of China under grant 2015CB351806, the National Natural Science Foundation of China under contract No. 61425025, No. 61390515 and No. 61421062, and Shenzhen Peacock Plan.

The rest of the paper is organized as follows. The related works are briefly discussed in section 2. Scene video coding techniques are introduced in section 3. Section 4 contains the experimental results of AVS2 scene video coding. The paper is concluded in section 5.

## 2 Related Works

Some research has been done to improve the compression ef-



▲ Figure 1. The architecture of AVS2 scene video coding.

iciency in scene videos.

One of the most direct solutions for surveillance and conference videos is the object-based coding. In the object-oriented analysis-synthesis coding method, each video was coded with motion and shape of objects, color information and prediction residuals. However, object-based coding has three main challenges: accurate foreground segmentation, low-cost object representation, and high-efficiency foreground residual coding [2].

In the traditional hybrid coding framework, hybrid block-based methods are used to encode each picture block by block. The main types of these methods include the following aspects: 1) Region-based coding and 2) Background prediction based coding. The former aimed at achieving better subjective quality of foreground regions with low coding complexity. Instead, with the assumption that in scene videos, there might be one background picture that remains unchanged for a long time, the second method improves the objective compression efficiency by utilizing one background picture as the reference for the following pictures.

However, there are some regions that may appear in the current frame but are covered by objects in the recent reference frames or the key frame. Thus, it is hard to compress the regions efficiently by using the key frame as the background. To address this problem, several background modeling based methods were proposed, for example, using the reconstructed pictures to model the background or utilizing the background picture that was modeled from the original input frames as the reference for more efficient background prediction.

### 3 AVS2 Scene Video Coding Techniques

As we know, the key to improve the coding performance efficiently of scene videos is reducing the background redundancy. AVS2 adopts the long-term reference technique and S picture to reduce the background redundancy to improve the coding performance efficiently [3].

#### 3.1 The Long-Term Reference Technique

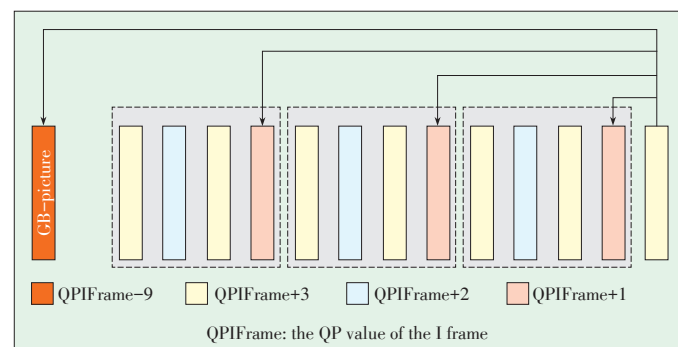
Traditionally, the current frame can only be inter-predicted by the previous frames in the group of pictures (GOP). Thus, the distance between the current picture and the reference frame is only relatively short, which means that the reference frame may not be able to provide abundant prediction in the background regions. In order to provide better reference for background regions, AVS2 adopts a long-term reference frame named background picture (GB picture) [1], [4].

As shown in **Fig. 2**, GB picture is a background picture where the whole picture is background regions, so the background regions of each subsequent inter-predicted frame can always find the matching regions in GB picture. When encoding the GB picture, only intra mode is utilized, and smaller Quantization Parameter (QP) is selected to obtain a high quality GB picture. When the P picture is uses the long-term refer-

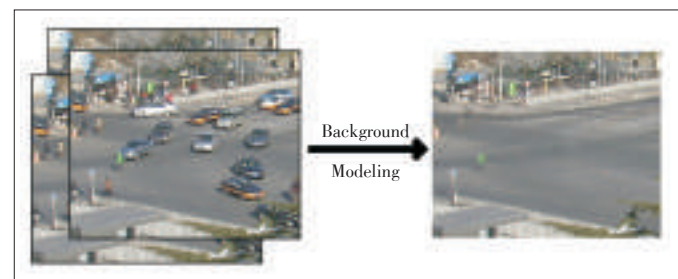
ence technique, the reference picture chain comprises general reference sequence and the long-term GB picture. Thus when the subsequent inter-predicted frames choose reference frames for their background regions, there is high possibility to select GB picture. Although the GB picture takes a lot of bits, more bitrate will be saved when the following frames refer GB picture because of the high quality of GB picture. As a result, the total performance becomes better.

Although the AVS2 standard does not limit the way a GB picture is generated, it chooses the segment-and-weight based running average (SWRA) to generate the GB picture in AVS2 Reference Design (RD). By weighting the frequent values more heavily in the averaging process, SWRA can generate pure background. The specific process is as follows. Technologically, SWRA divides the pixels at a position in the training pictures into temporal segments with their own mean values and weights and then calculates the running and weighted average result on the mean values of the segments. In the process, pixels in the same segment have the same background/foreground property, and the long segments are more heavily weighted. Experimental results [5] show that SWRA can achieve good performance yet without suffering a large memory cost and high computational complexity, which can meet the requirement of real-time transmission and storage for scene videos. An example of the constructed background frame and the training frames are shown in **Fig. 3**.

Once a GB picture is obtained, it is encoded, and the reconstructed picture is stored in the independent background memory and updated only if a new GB picture is selected or generated. The update mechanism guarantees the effectiveness of the



▲ **Figure 2.** The long-term reference technique.



▲ **Figure 3.** The training frames and the background frame.

## Introduction to AVS2 Scene Video Coding Techniques

Jiaying Yan, Siwei Dong, Yonghong Tian, and Tiejun Huang

GB picture.

### 3.2 S Picture for Random Access

To ensure random access ability, the picture at the random access point is decoded independent of the previous frames. In previous coding frame standards, I picture can be used as the random access point. However, the performance of I picture is not very well because it only adopts intra prediction and the performance of intra prediction is not equal to inter prediction. Along with GB picture and the long-term reference technique, another picture type called S picture is designed for balancing the coding performance and purpose of random access.

S picture is similar to P picture, which can only be predicted from a reconstructed GB picture and has no motion vector, so only three modes including Intra, Skip and 2N×2N are available in S picture. These characteristics make it possible for an S picture to be an ideal replacement for an I picture. Before the S picture is generated, the GB picture is first obtained to ensure the decoding independence of S picture. Zero motion vector in the S picture also makes sure that there is no need in consideration of the motion vector prediction (MVP). Thus the relative independence of the S picture makes sure it can be set as the random access point and can present better performance than I picture.

### 3.3 Improvement of Motion Vector Derivation

The BlockDistance is the distance between the current block and the reference block pointed by the motion vector, which is associated with the picture order count (POC) of reference picture. In the case of one reference with two motion vectors, such as F picture, one of the motion vectors is calculated by the other motion vector. When we introduce GB picture to AVS2, the problem comes. Because there is no POC existing in GB picture, the BlockDistance between current block and its reference block is unavailable if the reference block is from GB picture. In order to solve the problem, AVS2 provides a strategy in this situation. If one of the reference pictures is GB picture, the BlockDistance between current block and the block in GB picture is restricted to 1. By doing so, the motion vector derivation is available all the time no matter GB picture is involved in or not.

## 4 Performance Evaluation of AVS2 Scene Video Coding

## 4.1 Common Test Sequences and Conditions

There are five typical scene videos selected as the common test sequences [6], [7]. Three are 720×576 surveillance videos, and the other two are 1600×1200 ones (**Table 1**). From **Fig. 4**, these five surveillance videos cover different monitoring scenes, including bright and dusky lightness (BR/DU), large and small foreground (LF/SF), fast and slow motion (FM/SM).

**Table 1. The common test sequences of AVS2 scene video coding**

Resolution	FrameRate	Sequence	FramesToBeEncoded
720×576	30	Crossroad	600
		Office	
		Overbridge	
1600×1200	30	Intersection	600
		Mainroad	



▲ Figure 4. The common test sequences for scene video coding in AVS2.

To evaluate the coding performance of the scene video compression of AVS2 (RD 12.0.1 Scene), the latest released reference software for AVS2 keeping the scene video coding techniques disabled (RD 12.0.1 General) is used as the basic experimental platform. Here, our objective is to evaluate the improvement in efficiency and reduction in complexity that AVS2 scene video coding can achieve over AVS2 General.

Four configurations are adopted to perform the experiment [8]. They are:

- 1) Low delay (LD);
- 2) Random Access with B slices (RAB);
- 3) Random Access with F slices (RAF);
- 4) Random Access with P slices (RAP).

The F frame is a bidirectional reference frame. Unlike the B frame, one motion vector of the F frame is derived from the other motion vector.

**Table 2** shows the common test conditions of AVS2 scene video coding.

## 4.2 Performance Evaluation

The coding performance between RD 12.0.1 Scene and RD 12.0.1 General is shown in **Table 3**. According to the experimental result, RD 12.0.1 Scene reduces 24.33% (LD), 44.11% (RAB), 40.25% (RAF) and 40.56% (RAP) bitrates in average against RD 12.0.1 General on  $720 \times 576$  videos and 42.07% (LD), 39.24% (RAB), 38.36% (RAF) and 37.90% (RAP) on  $1600 \times 1200$  videos. Among the video sequences, Office and Intersection have large foreground objects and they are hard to generate clear background picture, so the coding performance

## Introduction to AVS2 Scene Video Coding Techniques

Jiaying Yan, Siwei Dong, Yonghong Tian, and Tiejun Huang

▼ Table 2. The common test conditions of AVS2 scene video coding

Parameter	LD	RAB	RAF	RAP
QPIFrame		27, 32, 38, 45		
QPPFrame		QPIFrame+1		
QPBFrame	-	QPIFrame+4	-	-
SeqHeaderPeriod	0	1	1	1
IntraPeriod	0	32	32	32
NumberBFrames	0	7	0	0
FrameSkip	0	7	0	0
BackgroundQP		QPIFrame-9		
BackgroundEnable		1		
FFRAMEEnable	1	1	1	0
ModelNumber		120		
BackgroundPeriod	900	112	900	900
LD: low delay		RAF: random access with F slices		
RAB: random access with B slices		RAP: random access with P slices		

▼ Table 3. The coding performance comparison between RD 12.0.1 Scene and RD 12.0.1 General

Resolution	Sequence	RD 12.0.1 Scene vs. RD 12.0.1 General (BD-Rate)			
		LD	RAB	RAF	RAP
720×576	Crossroad	-25.64%	-41.99%	-37.48%	-38.07%
	Office	-12.66%	-26.77%	-23.77%	-24.10%
	Overbridge	-34.10%	-63.58%	-59.50%	-59.51%
	Average	-24.13%	-44.11%	-40.25%	-40.56%
1600×1200	Intersection	-22.46%	-22.06%	-21.19%	-19.91%
	Mainroad	-61.68%	-56.42%	-55.52%	-55.90%
	Average	-42.07%	-39.24%	-38.36%	-37.90%
All	Average	-31.31%	-42.16%	-39.49%	-39.50%
BD: Bjøntegaard-Delta		RAB: random access with B slices			
RD: Reference Design		RAF: random access with F slices			
LD: low delay		RAP: random access with P slices			

is relatively lower than others. In average, RD 12.0.1 Scene can obtain 31.31% (LD), 42.16% (RAB), 39.49% (RAF) and 39.50% (RAP) bitrate savings on all common test sequences.

## 5 Conclusions

Based on the classic block-based hybrid video framework, AVS2 is the latest coding standard with efficient scene video coding techniques and is designed for high efficiency video coding of scene videos. This paper introduces several representative techniques adopted in AVS2, including the long-term reference technique and S picture.

By adopting the techniques of scene video coding mentioned above, AVS2 can gain 31.31% (LD), 42.16% (RAB), 39.49% (RAF) and 39.50% (RAP) BD-rate saving in coding efficiency on scene videos. The excellent coding performance of AVS2 in

scene videos coding will bring a bright prospect in video coding research and industrial fields.

## References

- [1] L. Zhao, S. Dong, P. Xing, and X. Zhang, "AVS2 surveillance video coding platform," AVS M3221, Dec. 2013.
- [2] X. Zhang, Y. Tian, T. Huang, S. Dong, and W. Gao, "Optimizing the hierarchical prediction and coding in hevc for surveillance and conference videos with background modeling," *IEEE Transaction on Image Processing*, vol. 23, no.10, pp. 4511–4526, Oct. 2014. doi: 10.1109/TIP.2014.2352036.
- [3] F. Liang, "Information technology—advanced media coding part2: video (FCD4)," AVS N2216, Sept. 2015.
- [4] R. Wang, Z. Ren, H. Wang, "Background-predictive picture for video coding," AVS M2189, Dec. 2007.
- [5] X. Zhang, Y. Tian, T. Huang, and W. Gao, "Low-complexity and high-efficiency background modelling for surveillance video coding," in *Proc. IEEE International Conference on Visual Communication and Image Processing*, San Diego, USA, Nov. 2012, pp. 1–6. doi: 10.1109/VCIP.2012.6410796.
- [6] S. Dong, L. Zhao, "AVS2 surveillance test sequences," AVS M3168, Sept. 2013.
- [7] L. Yu, "Meeting summary of AVS2 video coding subgroup," AVS N1998, Sept. 2013.
- [8] X. Zheng, "Common test conditions of AVS2-P2 surveillance profile," AVS N2217, Sept. 2015.

Manuscript received: 2015-11-25

## Biographies

**Jiaying Yan** (yanjiaying@pku.edu.cn) received the BS degree from Beijing Institute of Technology, China in 2014. He is currently pursuing the MS degree with the School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, China. His research interests include surveillance video coding and multimedia learning.

**Siwei Dong** (swdong@pku.edu.cn) received the B.S. degree from Chongqing University, China in 2012. He is currently pursuing the PhD degree with the School of Electronics Engineering and Computer Science, Peking University, China. His research interests include video coding and multimedia learning.

**Yonghong Tian** (yhtian@pku.edu.cn) is currently a professor with the National Engineering Laboratory for Video Technology, School of Electronics Engineering and Computer Science, Peking University, China. He received the PhD degree from the Institute of Computing Technology, Chinese Academy of Sciences, China in 2005, and was also a visiting scientist at Department of Computer Science/Engineering, University of Minnesota, USA from November 2009 to July 2010. His research interests include machine learning, computer vision, video analysis and coding, and multimedia big data. He is the author or coauthor of over 110 technical articles in refereed journals and Conferences. Dr. Tian is currently an associate editor of *IEEE Transactions on Multimedia*, a young associate editor of the *Frontiers of Computer Science*, and a member of the IEEE TCMC-TCSEM Joint Executive Committee in Asia (JECA). He was the recipient of the Second Prize of National Science and Technology Progress Awards in 2010, the best performer in the TRECVID content-based copy detection (CCD) task (2010–2011), the top performer in the TRECVID retrospective surveillance event detection (SED) task (2009–2012), and the winner of the WikipediaMM task in ImageCLEF 2008. He is a senior member of IEEE and a member of ACM.

**Tiejun Huang** (tjhuang@pku.edu.cn) is a professor with the School of Electronic Engineering and Computer Science, the chair of Department of Computer Science and the director of the Institute for Digital Media Technology, Peking University, China. His research areas include video coding and image understanding, especially neural coding inspired information coding theory in last years. He received the PhD degree in pattern recognition and intelligent system from the Huazhong (Central China) University of Science and Technology in 1998, and the master's and bachelor's degrees in computer science from the Wuhan University of Technology in 1995 and 1992, respectively. Professor Huang received the National Science Fund for Distinguished Young Scholars of China in 2014. He is a member of the Board of the Chinese Institute of Electronics, the Board of Directors for Digital Media Project and the Advisory Board of IEEE Computing Now.



# From CIA to PDR: A Top-Down Survey of SDN Security for Cloud DCN

Zhi Liu<sup>1, 2</sup>, Xiang Wang<sup>1, 2</sup>, and Jun Li<sup>1, 3</sup>

(1. Research Institute of Information Technology, Tsinghua University, Beijing 100084, China;

2. Department of Automation, Tsinghua University, Beijing 100084, China;

3. Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China)

## 1 Introduction

Information technology has come a long way — from mainframes to personal computing and on to mobile computing. Now we are embracing cloud computing that was previously called utility computing or grid computing. In this fascinating transition, datacenters are similar to the mainframes of the old days, and mobile devices are like the old terminals, only much smarter and not tethered.

Traditional datacenters usually host proprietary services backed by a number of static and tightly coupled applications. Traditional datacenter networks (DCNs) mainly deal with large volumes of north-south traffic and usually have three layers (Fig. 1a). The access layer provides the connectivity for servers and storage facilities, normally through top-of-rack (ToR) switches. The aggregation layer mediates the access layer to the core layer, which in turn interfaces to the Internet. As the cloud evolves towards virtualization and multi-tenancy, this architecture often lacks elasticity and suffers from vendor lock-in [1].

Modern cloud datacenters support a variety of heterogeneous services for multiple tenants simultaneously. These datacenters are commonly built with a two-tier DCN (Fig. 1b). Tenants can deploy their own services on the shared infrastructure and pay-as-they-go. Several software-defined datacenter (SD-DC) solutions have been proposed so that capacity can be expanded using infrastructure multiplexing and all tenant systems can be managed in an efficient, automatic manner.

Making datacenter services public instead of proprietary significantly increases infrastructure utilization and drastically affects the DCN design. Virtual machines (VMs) are frequently brought up, shut down, and even migrated across datacenters. Moreover, VMs of the same tenant may interconnect across

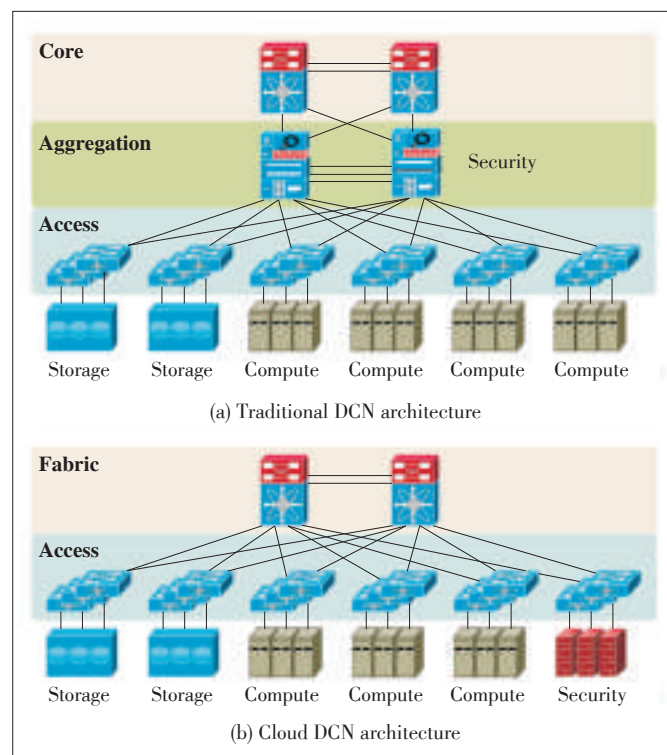
## Abstract

By extracting the control plane from the data plane, SDN enables unprecedented flexibility for future network architectures and quickly changes the landscape of the networking industry. Although the maturity of commonly accepted SDN security practices is the key to the proliferation of cloud DCN, SDN security research is still in its infancy. This paper gives a top-down survey of the approaches in this area, discussing security challenges and opportunities of software-defined datacenter networking for cloud computing. It leverages the well-known confidentiality-integrity-availability (CIA) matrix and protection-detection-reaction (PDR) model to give an overview of current security threats and security measures. It also discusses promising research directions in this field.

## Keywords

SDN security; cloud DCN; CIA; PDR

multiple physical servers, and VMs of different tenants may share the same physical server. These complex scenarios make it very difficult to guarantee service-level agreements (SLAs)



▲ Figure 1. Evolution of DCN architecture.



for each and every tenant.

In the cloud era, novel technologies are required to cope with emerging DCN security challenges [2]. Such technologies include topology-independent service assignment and policy enforcement, flow-based (rather than packet-based) processing, and awareness of virtualization and multi-tenancy. From many industrial surveys, we see that security concerns are still an obstacle to the proliferation of cloud computing [3].

Software-defined networking (SDN) is central to addressing complex network management and security issues. It decouples the control plane from the data plane by extracting the mostly autonomous embedded controllers from traditional network elements. The virtually centralized SDN control plane leverages its global knowledge of network topology and status, and acts as a network operating system. This enables a network development and operation (DevOps) team to program network services via open and standard application programming interfaces (APIs) such as OpenFlow. This also instigates the rise of white boxes, as opposed to closed proprietary products of a few dominant vendors.

Because SDN is not yet mature, cloud DCN security is in its infancy. Cloud DCN security is a hot research topic and there is no consensus on it yet. Standardization and industrial application of cloud DCN security is still at a very early stage. This paper focuses on the challenges and opportunities related to cloud DCN. We provide a top-down survey of recent approaches to SDN security and employ the confidentiality-integrity-availability (CIA) matrix [4] and protection-detection-response (PDR) model [5] for analyzing security threats and measures. Section 2 reviews related work. Section 3 discusses DCN building blocks and corresponding security demands. Section 4 and section 5 summarize security threats and security measures, respectively. Section 6 concludes the paper.

## 2 Previous Work

Although SDN and network function virtualization (NFV) are very recent trends in networking, several comprehensive surveys of related security research and technologies have already been published [6]–[9]. Some are even updated from time to time to reflect the fast progress in this area. Existing surveys have different perspectives on SDN security. Some distinguish between research on protecting the network and research on providing security as a service, i.e., secure SDN (security of SDN) and SDN security (security by SDN) [6], [7]. Others analyze and summarize SDN security technologies in different target environments [8] or according to types of middlebox functions [9].

In [7], the authors review SDN characteristics and present a survey of security analysis and potential threats in SDN. They then describe a holistic approach to designing the security architecture required by SDN. Their summary of the problems and solutions for each of the main threats to SDN is helpful for

an overall understanding of SDN security advances. The authors conclude that, evidenced by the commercially available applications, work on leveraging SDN to increase network security is more mature than the solutions addressing the security issues inherited or introduced by SDN.

In [8], the authors give an overview of existing research on SDN security, focusing on an analysis of security threats and potential damage. Such threats include spoofing, tampering, repudiation, information disclosure, denial-of-service (DoS), and elevation of privilege. The authors also discuss SDN security measures, such as firewall, intrusion detection system (IDS) (or intrusion prevention system, IPS), policy management, monitoring, auditing, privacy protection, and others controls to threats in specific networking scenarios. A comprehensive list of references categorized into different SDN security functionalities is provided.

This paper takes a more fundamental and focused point of view from the perspective of practical conditions. We first partition the cloud DCN into intra-DCN, access-DCN, and inter-DCN, and differentiate the unique properties of them. Then, we analyze the changing attributes of the traditional PDR model from the perspective of CIA matrix.

## 3 The Three Networks

In a traditional DCN, there are various middleboxes that provide rich network services in addition to basic connectivity offered by forwarding devices, such as switches and routers. Firewalls, IDS/IPS, and other security middleboxes are normally deployed at the aggregation layer to inspect and steer network traffic. In this outdated model, policy enforcement is closely coupled with actual reachability; therefore, the middleboxes have to sit on the physical packet path, causing administration difficulties and performance bottlenecks [2].

Leveraging SDN, cloud DCN relies on a flat architecture to achieve better elasticity and is designed for cost efficiency and performance enhancement. In this new model, especially in public cloud DCN with pervasive multi-tenancy and high resource utilization, north-south traffic gives way to east-west traffic [10]. The hierarchical partition of the DCN is no longer valid, and DCN building blocks can be categorized according to functional characteristics, such as intra-DCN, inter-DCN, and access-DCN [10] (**Fig. 2**).

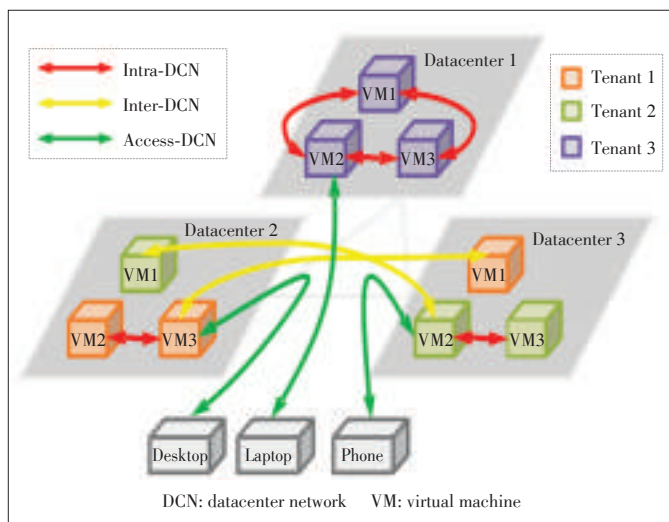
### 3.1 Intra-Datacenter Network

Intra-DCN is the network of resources inside the datacenter. The intra-DCN connects all IT elements together to create clouds for tenants. With virtualization and multi-tenancy, the clear network boundaries between traditional security zones of networks usually disappear; thus, network security policies are now enforced on dynamically distributed network security functions [11].

The correctness and efficiency of security policy deployment

## From CIA to PDR: A Top-Down Survey of SDN Security for Cloud DCN

Zhi Liu, Xiang Wang, and Jun Li



▲ Figure 2. Three networks of cloud DCN.

depends on the controller's real-time awareness of network topology, service status, and traffic pattern. Several approaches to providing a wide variety of security functionalities for intra-DCN and adapting to network changes have been proposed. VXLAN [12] and a few other encapsulation protocols are deployed for network virtualization to isolate traffic of different tenants or subnets. Service function chaining (SFC) [13] has been proposed to orchestrate multiple middleboxes of the same or different functions. Micro-segmentation provides middlebox functions within L2 networks and delivers fine-grained network security. In OpenStack, the most promising open-source cloud platform, neutron network service program also incubates firewall-as-a-service (FWaaS), VPN-as-a-service (VPNaaS), IDS-as-a-service (IDSaaS), and load-balancing-as-a-service (LBaaS) projects for security service provision within cloud datacenters.

In existing works on intra-DCN security, the focus is on providing security capacity and functions with agility and elasticity.

### 3.2 Access-Datacenter Network

Access-DCN is the network of clients outside datacenters that provide direct and pervasive connectivity for users so that they can access applications running in the cloud.

Distributed denial-of-service (DDoS) is one of the most hotly discussed topics related to access interfaces of cloud datacenters. There has been some recent advancement on the application delivery controller (ADC) and web application firewall (WAF). Other work has also been done on mobile access and more application-specific areas. Existing network security devices, such as unified threat management (UTM) and next-generation firewall (NGFW), can also provide high performance at this location, including hardware accelerations.

Because the access points of all tenants are connected to the Internet, which shares the same IP address space, tenants can

take full advantage of the security hardware resources to complete common security inspections. In summary, solutions of access-DCN security mainly focus on optimizing security inspections.

### 3.3 Inter-Datacenter Network

Inter-DCN is the network of clouds for federation networking between public and private cloud datacenters or optimizing network resources between multiple datacenter sites.

Google's B4 [14] is the most influential achievement in inter-datacenter networking. Microsoft's software-driven WAN [15] is also constructed for peak load shifting. There has not been a lot of security R&D on this front, mostly because mature virtual private network (VPN) technologies already satisfy the basic security requirements of cloud providers.

The rest of this paper is mainly focused on intra-DCN, which is the focal point of network security and advancement.

## 4 The Three Threats

Network security threats are becoming more sophisticated and powerful. Advanced persistent threat (APT) uses blended hacking schemes to penetrate a network and compromise the target systems. Recent DDoS attacks have reached 400 Gbps aggregated network traffic volume, and the number of attacks over 100 Gbps has greatly increased [16]. Network security threats all basically boil down to interception, modification, interruption, and fabrication. The fundamental security matrix is still CIA, although authenticity, non-repudiation, and other security mechanisms are equally important.

### 4.1 Threat to Confidentiality

In cloud datacenters, confidentiality may be ensured by access control list (ACL) and cryptographic solutions. However, the fundamental challenge lies in tenant isolation. For intra-DCN, this means tenant traffic isolation: one tenant should not be able to send or receive network packets to or from another tenant unless explicitly permitted by the security policy. Tenant isolation is a key feature supported by the SDN virtual networking PaaS.

PortLand [1] is an example of the design and implementation of a non-blocking network fabric for virtualized datacenters. Multi-tenancy and tenant isolation are achieved by changing the processing logic of access switches with the rewriting of hierarchical pseudo MAC addresses. NetLord [17] proposes an encapsulation scheme for overlay network virtualization. It can be deployed on existing networking devices without any modification and enables different tenants to share the same L2/L3 address spaces. NetLord also has very good scalability.

NVP [18] describes the overall design of network virtualization platform, including both data plane and control plane. It leverages Open vSwitch and packet encapsulation to implement the overlay network virtualization, and designs a datalog-

based declaration language to define and implement network policy. LiveCloud [19] further addresses the integration of hardware networking devices in clouds. It uses both hardware and software switches to compose the access layer for various resources.

Reviewing these existing approaches, it can be observed that traffic is almost always isolated at the network edge, where the bulk of computing resources can be used for complex processing logic. At the same time, this requires dynamic policy coordination and deployment for on-demand stateful inspection, such as NFV-ed firewall, to ensure that policies are globally correct and locally conflict-free.

#### 4.2 Threat to Integrity

In terms of integrity in the broader perspective, deep inspection prevents intrusion and/or extrusion and is the most critical demand [20], including NFV-ed IDS/IPS and data leakage prevention (DLP).

Player [2] introduces a policy-aware switching layer for deployment of middleboxes. This approach removes middleboxes from traffic paths and steers traffic to traverse these devices in a user-defined sequence. It essentially decouples network policies from physical topology, which introduces much more flexibility into middlebox deployment. SIMPLE [21] addresses the same problem but also solves the problems of traffic routing loops and the negative effects of packet modification. It also takes into consideration routing and load balancing given switch constraints. A reliable solution for dynamic middlebox actions, FlowTags [22] designs a tagging scheme that exposes the internal mapping of flows before and after middlebox processing. The introduced tags can be recognized and leveraged by SDN switches to compose service chains.

On the control plane level, Stratos [23] proposes a framework for middlebox orchestration according to workload variation. Tackling the closed middlebox implementation in Stratos, OpenNF [24] abstracts the middlebox API and designs a series of APIs for middlebox configuration and notification. These APIs can be used to coordinate the state control of both middleboxes and forwarding devices. SDSA [25] introduces a dedicated security controller for security-related functionalities, such as security device management, security policy deployment, and security event monitoring. The security controller also cooperates with the network controller to obtain a global view and enforce security policies such as ACL. Considering topology changes caused by VM migration and dynamic resource relocation, real-time security capacity redistribution and policy instance update are vitally important.

#### 4.3 Threat to Availability

In terms of availability, most security efforts are directed towards DoS/DDoS mitigation. To counter attacks and prevent service unavailability, security middleboxes and policies are often deployed dynamically on these middleboxes. DFence [26]

dynamically instantiates DDoS mitigation middleboxes, intercepts suspicious network traffic, and filters attacking traffic. A dynamic throttling method was also proposed in [27] to prevent DoS attack. With this method, flows originating from the same client are limited when the request rate from the client exceeds a dynamically determined threshold. Pushback [28] has a cooperative mechanism to mitigate DDoS attacks. The rate of upstream devices is limited when a DDoS attack occurs so that the attacking traffic is blocked near its entry point.

Availability security threats have diverse mechanisms for every specific scenario, which means the identification of suspicious traffic patterns (defined by security operators and expressed in the security policy) is very important. Thus, the management of security policies is central to intra-DCN security. Management of security policies includes policy definition [29], [30], policy compilation [31], [32], policy assignment [33], [34], policy optimization [35], [36], policy deployment [37], [38], and policy lookup [39], [40]. Some research has described several roadmaps ahead, but so far no consensus has been reached.

### 5 The Three Stages

Security is mostly a defensive practice that takes charge of policy enforcement. From the perspective of control theory, articulate system design is required to meet application requirements, where sensors and actuators are versatile for real time response, and feedback is essential to constantly adapt the situational changes and improve control quality. Many security approaches targeting the SDN-based cloud DCN have been proposed and can be evaluated in the well-known PDR lifecycle model.

#### 5.1 Protection Stage

In the protection (or planning) stage, the key to intra-DCN security is to design a suitable architecture that both satisfies the security management requirements and is future-proof to a certain extent.

Unlike traditional DCN, SDN has a global view of the cloud DCN, and thus enables security mechanisms to be deployed in a distributed and dynamic manner. Two aspects need to be weighed in this phase: where to place security functions and how to manage security policies.

Regarding the placement of security functions, SDN and NFV devices are orthogonal [41]. **Table 1** shows the main differences between SDN and NFV. SDN focuses on network forwarding, mainly for traffic delivery. It performs stateless processing of L2-L3 network traffic at the packet level according to network topology. By contrast, the basic responsibility of NFV is network monitoring, and it is also responsible for security, measurement, and optimization. NFV conducts stateful and deep inspection of L4-L7 network traffic at the flow level according to resources and policies.

## From CIA to PDR: A Top-Down Survey of SDN Security for Cloud DCN

Zhi Liu, Xiang Wang, and Jun Li

▼Table 1. Orthogonality of SDN and NFV

	Forwarding	Monitoring
Task	Delivery	Security, measurement, optimization
Logical object	Packet	Flow
Physical object	L2-L3, Header	L2-L7, Header + Payload
Basis	Topology	Resource, policy
State	Stateless	Stateful
Manner	Local autonomy	Global governance
Device	NIC, hub, switch, router	Middlebox
Algorithm	Routing origination, Routing lookup	Packet classification, pattern matching, AppID, traffic management
	NIC: Network Interface Card	AppID: Application Identification

Conventionally, SDN and NFV devices are managed by different administrators. Tualatin [11] is designed according to orthogonal principles and provides efficient security in a cloud datacenter. Networking devices and security devices are separately managed by their corresponding controllers (Fig. 3). Considering both flexibility and performance, Tualatin decouples the security scenarios into intra-VN, inter-VN, and access-VN and uses hardware and software co-design to meet different security requirements.

There have also been proposals of pushing all middleboxes, mostly network security functions, to the edge of intra-DCN [42] or implementing security inspected in off-path control plane [43]. However, the authors of this paper do not believe this will solve the problem all together.

Security policies can be enforced with changing [44] or re-

specting [33] forwarding policies. Security policy enforcement combined with forwarding policies can easily introduce performance impact on the data plane, while security policy enforcement based on forwarding policy has clear design boundary and thus simplifies control plane structure.

### 5.2 Detection Stage

In the detection (or runtime) stage, network security functions are used to discover and defend security attacks.

In Fig. 3, intra-VN security depends on traffic statistics generated by NetFlow on software switches to enable heavy-load security inspections. Both ACL and QoS policies are deployed on software switches. For inter-VN security, Tualatin chains multiple security services within a standalone virtual network. Tualatin introduces a security workload scheduler for load balancing and function composition and exposes fine-grain APIs for flow slice to support micro-segmentation. For access-VN, hardware UTM or NGFW can be leveraged for common security inspections for multiple tenants. This helps with the sharing of computing resources of security devices.

To efficiently implement these detection engines, virtualized middleboxes need to be redesigned in a consolidated way. RouteBricks [45] reveals the curtain of high-speed packet processing on commodity servers. CoMB [46] consolidates middlebox functions and re-implements them on an X86 platform. These works demonstrate the possibility of high-performance middleboxes on commodity servers, which lays the foundation for NFV. OpenGate [47] proposes the architecture for distributed middlebox processing. It takes full advantage of different hardware platforms to tackle the challenges of L2-L3 and L4-L7 processing, which helps to optimize middlebox performance.

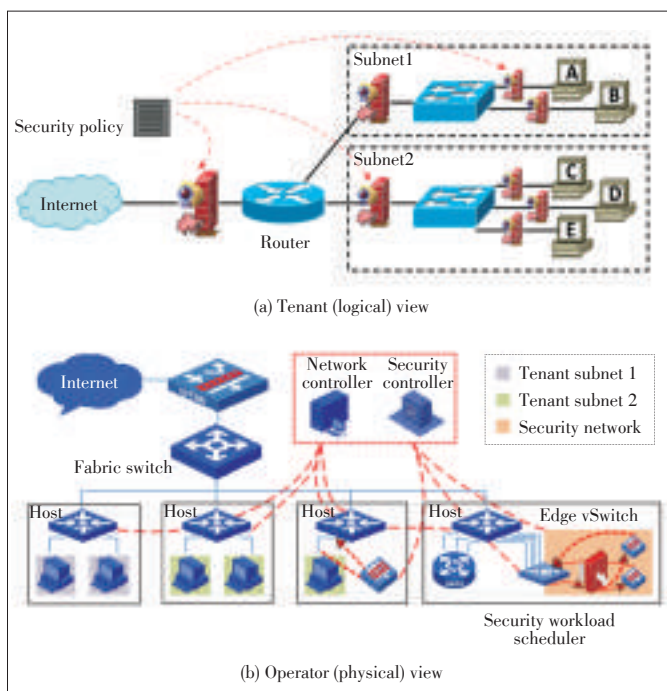
For all these hardware accelerated or software virtualized functions to cooperate effectively and achieve high performance, major breakthroughs in policy management technology is necessary.

### 5.3 Response Stage

In the response (or feedback) stage, security events, action results, clues of potential threats, statistical and behavioral anomalies are collected. The collected information is analyzed with special tools, including machine learning and big data, to find new threat signatures or models [48], previously unknown vulnerabilities [49], and ways to improve security back to the protection stage. Within industry, security information and event management (SIEM) advancements can definitely be leveraged on this front [50].

## 6 Conclusion

Modern DCN for cloud computing has made great progress in terms of architecture evolution, and now SDN and NFV are leading the way forward. Therefore, SDN security is critical for



▲Figure 3. Security service in a cloud datacenter.



the proliferation of multifarious cloud services.

Despite the extrinsic nature of various emerging threats—especially those introduced by virtualization and multi-tenancy—the essence of network security is still unchanged. Beginning with the well-known PDR model, this paper has discussed the latest threats categorized by the CIA matrix as well as network security advancements.

In the area of intra-DCN security, this paper emphasizes the central role of security policies in the evolution of novel security mechanisms, including network virtualization and isolation, intrusion and extrusion prevention, and attack defense and mitigation. From security architecture to particular algorithms, from theory to practice, from academia to industry, there have been more and more proposals and developments around different aspects of policy management, such as definition, compilation, assignment, optimization, deployment and lookup.

Besides the management of security policy, other notable challenges and opportunities have unveiled promising directions in the green field of DCN security. We argue that there is yet no sign of framework consensus or approach convergence in the near future for SDN based cloud DCN security, and we expect key developments in distributed policy, service chaining, as well as visualization and troubleshooting tools.

## References

- [1] R. N. Mysore, A. Pamboris, N. Farrington, *et al.*, "PortLand: a scalable fault-tolerant layer 2 data center network fabric," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4, pp. 39-50, 2009. doi: 10.1145/1592568.1592575.
- [2] D. A. Joseph, A. Tavakoli, and I. Stoica, "A policy-aware switching layer for data centers," *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 51-62, 2008. doi: 10.1145/1402958.1402966.
- [3] Forbes. Predicting Enterprise Cloud Computing Growth [Online]. Available: <http://www.forbes.com/sites/louisclumbus/2013/09/04/predicting-enterprise-cloud-computing-growth/>
- [4] U.S. Government Publishing Office. Public Printing and Documents - Coordination of Federal Information Policy - Information Security - Definitions [Online]. Available: <http://www.gpo.gov/fdsys/granule/USCODE-2011-title44/USCODE-2011-title44-chap35-subchapIII-sec3542/content-detail.html>
- [5] W. Schwartau. "Time based security." New York, USA: Interact Press, 1999.
- [6] S. T. Ali, V. Sivaraman, A. Radford, and S. Jha, "A survey of securing networks using software defined networking," *IEEE Transactions on Reliability*, vol. 64, no. 3, pp. 1086-1097, Sept. 2015. doi: 10.1109/tr.2015.2421391.
- [7] S. Scott-Hayward, S. Natarajan, and S. Sezer, "A survey of security in software defined networks," *IEEE Communications Surveys & Tutorials*, vol. PP, no. 99, p. 1, Jul. 2015. doi: 10.1109/comst.2015.2453114.
- [8] I. Alsmadi and D. Xu, "Security of software defined networks: a survey," *Computers & Security*, vol. 53, pp. 79 - 108, Sep. 2015. doi: 10.1016/j.cose.2015.05.006.
- [9] J. François, L. Dolberg, O. Festor, and T. Engel, "Network security through software defined networking: a survey," in *Proc. Conference on Principles, Systems and Applications of IP Telecommunications*, Chicago, USA, 2014, p. 6. doi: 10.1145/2670386.2670390.
- [10] Cisco Systems. Cisco Global Cloud Index: Forecast and Methodology, 2014-2019 [Online]. Available: [http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud\\_Index\\_White\\_Paper.html](http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html)
- [11] X. Wang, Z. Liu, B. Yang, Y. Qi, and J. Li, "Tualatin: towards network security service provision in cloud datacenters," in *IEEE 23rd International Conference on Computer Communication and Networks (ICCCN)*, Shanghai, China, 2014, pp. 1-8. doi: 10.1109/icccn.2014.6911782.
- [12] *Virtual Extensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks*, IETF RFC 7348, Aug. 2014.
- [13] *Service Function Chaining (SFC) Architecture*, IETF RFC 7665, Oct. 2015.
- [14] S. Jain, A. Kumar, S. Mandal, *et al.*, "B4: experience with a globally-deployed software defined wan," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 3-14, 2013. doi: 10.1145/2486001.2486019.
- [15] C. Hong, S. Kandula, R. Mahajan, *et al.*, "Achieving high utilization with software-driven WAN," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 15-26, 2013. doi: 10.1145/2486001.2486012.
- [16] Akamai. Q2 2015 State of the Internet—Security Report [Online]. Available: <https://www.stateoftheinternet.com/resources-cloud-security-2015-q2-web-security-report.html>
- [17] J. Mudigonda, P. Yalagandula, J. C. Mogul, B. Stiekes, and Y. Pouffary, "Net-Lord: a scalable multi-tenant network architecture for virtualized datacenters," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 62-73, 2011. doi: 10.1145/2018436.2018444.
- [18] T. Koponen, K. Amidon, P. Baland, *et al.*, "Network virtualization in multi-tenant datacenters," in *Proc. 11th USENIX Symposium on Networked Systems Design and Implementation*, Seattle, USA, Apr. 2014.
- [19] X. Wang, Z. Liu, Y. Qi, and J. Li, "LiveCloud: a lucid orchestrator for cloud datacenters," in *Proc. IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom)*, Los Alamitos, USA, pp. 341-348, Dec. 2012. doi: 10.1109/cloudcom.2012.6427544.
- [20] A. Bremner-Barr, Y. Harchol, D. Hay, and Y. Koral, "Deep packet inspection as a service," in *Proc. 10th ACM International Conference on Emerging Networking Experiments and Technologies*, Sydney, Australia, pp. 271-282, 2014.
- [21] Z. A. Qazi, C. Tu, L. Chiang, *et al.*, "SIMPLE-flying middlebox policy enforcement using SDN," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 27-38, 2013. doi: 10.1145/2486001.2486022.
- [22] S. Fayazbakhsh, L. Chiang, V. Sekar, M. Yu, and J. Mogul, "Enforcing network-wide policies in the presence of dynamic middlebox actions using FlowTags," in *Proc. 11th USENIX Symposium on Networked Systems Design and Implementation*, Seattle, USA, Apr. 2014, pp. 533-546.
- [23] A. Gember, A. Krishnamurthy, S. St. John, *et al.*, "Stratos: a network-aware orchestration layer for middleboxes in the cloud," University of Wisconsin-Madison, Madison, USA, Tech. Rep., 2013.
- [24] A. Gember, R. Viswanathan, C. Prakash, *et al.*, "OpenNF: enabling innovation in network function control," in *Proc. ACM Conference on SIGCOMM*, Chicago, USA, pp. 163-174, 2014. doi: 10.1145/2619239.2626313.
- [25] W. Liu, X. Qiu, P. Chen, *et al.*, "SDSA: a programmable software defined security platform," in *Proc. International Conference on Cloud Computing Research and Innovation*, Biopolis, Singapore, Oct. 2014, pp. 101-106.
- [26] A. Mahimkar, J. Dange, V. Shmatikov, H. Vin, and Y. Zhang, "dFence: transparent network-based denial of service mitigation," *Proc. 4th USENIX Symposium on Networked Systems Design and Implementation*, Cambridge, USA, Apr. 2007, pp. 327-340.
- [27] JE Belissent, "Method and apparatus for preventing a denial of service (DoS) attack by selectively throttling TCP/IP requests," U.S. Patent No. 6,789,203. 7, Sep. 2004.
- [28] J. Ioannidis and S. M. Bellovin, "Pushback: router-based defense against DDOS attacks," in *Proc. Network and Distributed System Security (NDSS) Symposium*, San Diego, USA, Feb. 2002. doi: 10.5353/th\_b3017330.
- [29] T. L. Hinrichs, N. Gude, M. Casado, J. C. Mitchell, and S. Shenker, "Practical declarative network manage," in *Proc. 1st ACM SIGCOMM Workshop on Research on Enterprise Networking*, Barcelona, Spain, Aug. 2009, pp. 1-10. doi: 10.1007/978-3-540-92995-6\_5.
- [30] C. Prakash, J. Lee, Y. Turner, *et al.*, "PGA: using graphs to express and automatically reconcile network policies," in *Proc. ACM Conference on Special Interest Group on Data Communication*, London, UK, Aug. 2015, pp. 29-42. doi: 10.1145/2785956.2787506.
- [31] N. Foster, R. Harrison, M. J. Freedman, *et al.*, "Frenetic: a network programming language," *ACM SIGPLAN Notices*, vol. 46, no. 9, pp. 279-291, 2011. doi: 10.1145/2034773.2034812.
- [32] C. Monsanto, J. Reich, N. Foster, J. Rexford, and D. Walker, "Composing software-defined networks," in *Proc. 10th USENIX Symposium on Networked Systems Design and Implementation*, Lombard, USA, Apr. 2013. doi: 10.1016/b978-0-12-416675-2.00014-0.
- [33] N. Kang, Z. Liu, J. Rexford, and D. Walker, "Optimizing the 'one big switch' abstraction in software-defined networks," in *Proc. Ninth ACM Conference on Emerging Networking Experiments and Technologies*, Santa Barbara, USA, Dec. 2013, pp. 13-24. doi: 10.1145/2535372.2535373.
- [34] X. Wang, W. Shi, Y. Xiang, and J. Li, "Efficient network security policy enforcement with policy space analysis," *IEEE/ACM Transactions on Networking*, 2016. doi: 10.1109/tnet.2015.2502402.
- [35] A. R. Curtis, J. C. Mogul, J. Tourrilhes, *et al.*, "DevoFlow: scaling flow manage-



## From CIA to PDR: A Top-Down Survey of SDN Security for Cloud DCN

Zhi Liu, Xiang Wang, and Jun Li

- ment for high-performance networks," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 254-265, 2011. doi: 10.1145/2018436.2018466.
- [36] P. Kazemian, G. Varghese, and N. McKeown, "Header space analysis: static checking for networks," in *Proc. 9th USENIX Symposium on Networked Systems Design and Implementation*, San Jose, USA, Apr. 2012, pp. 113-126.
- [37] M. Reitblatt, N. Foster, J. Rexford, C. Schlesinger, and D. Walker, "Abstractions for network update," in *Proc. ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, Helsinki, Finland, Aug. 2012, pp. 323-334. doi: 10.1145/2377677.2377748.
- [38] W. Zhou, D. Jin, J. Croft, M. Caesar, and P. Godfrey, "Enforcing customizable consistency properties in software-defined networks," in *Proc. 12th USENIX Symposium on Networked Systems Design and Implementation*, Oakland, USA, Apr. 2015, pp. 73-85.
- [39] B. Vamanan, G. Voskuilen, and T. Vijaykumar, "EffiCuts: optimizing packet classification for memory and throughput," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 207 - 218, 2011. doi: 10.1145/1851182.1851208.
- [40] Y. Qi, L. Xu, B. Yang, Y. Xue, and J. Li, "Packet classification algorithms: from theory to practice," in *Proc. 28th Conference on Computer Communications*, Rio de Janeiro, Brazil, Apr. 2009, pp. 648 - 656. doi: 10.1109/incom.2009.5061972.
- [41] J. McCauley, A. Panda, M. Casado, T. Koponen, and S. Shenker, "Extending SDN to large-scale networks," Open Networking Summit, Research Track, Santa Clara, USA, 2013.
- [42] S. Ratnasamy and S. Shenker. Quick Overview of SDN/NFV Research at Berkeley [Online]. Available: <http://onrc.stanford.edu/protected%20files/Day1/6.%20Overview%20of%20SDNv2%20Architecture%20and%20Related%20Efforts.pdf>
- [43] S. Shin, P. Porras, V. Yegneswaran, *et al.*, "FRESCO: modular composable security services for software-defined networks," in *Proc. 2014 Workshop on Security of Emerging Networking Technologies*, San Diego, USA. doi: 10.14722/sent.2014.23006.
- [44] M. Yu, J. Rexford, M. J. Freedman, and J. Wang, "Scalable flow-based networking with DIFANE," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 351-362, 2011. doi: 10.1145/1851182.1851224.
- [45] M. Dobrescu, N. Egi, K. Argyraki, *et al.*, "RouteBricks: exploiting parallelism to scale software routers," in *Proc. ACM SIGOPS 22nd Symposium on Operating Systems Principles*, Big Sky, USA, 2009, pp. 15 - 28. doi: 10.1145/1629575.1629578.
- [46] V. Sekar, N. Egi, S. Ratnasamy, M. Reiter, and G. Shi, "Design and implementation of a consolidated middlebox architecture," in *Proc. 9th USENIX Symposium on Networked Systems Design and Implementation*, San Jose, USA, Apr. 2012, pp. 24-24.
- [47] Y. Qi, F. He, X. Wang, *et al.*, "OpenGate: towards an open network services gateway," *Computer Communications*, vol. 34, no. 2, pp. 200-208, 2011.
- [48] M. V. Mahoney and P. K. Chan, "Learning nonstationary models of normal network traffic for detecting novel attacks," in *Proc. Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Canada, 2002, pp. 376-385. doi: 10.1145/775047.775102.
- [49] W. Fan, M. Miller, S. Stolfo, W. Lee, and P. Chan, "Using artificial anomalies to detect unknown and known network intrusions," *Knowledge and Information Systems*, vol. 6, no. 5, 2004, pp. 507-527. doi: 10.1109/ikdm.2001.989509.
- [50] N. B. Anuar, M. Papadaki, S. Furnell, and N. Clarke, "An investigation and survey of response options for intrusion response systems (IRSs)," in *IEEE Information Security for South Africa*, Johannesburg, South Africa, 2010, pp. 1-8. doi: 10.1109/issa.2010.5588654.

Manuscript received: 2015-12-01

## Biographies

**Zhi Liu** (zhi-liu12@mails.tsinghua.edu.cn) is currently a PhD candidate at Department of Automation, Tsinghua University, China. He received his BS degree from Department of Automation, Tsinghua University in 2012. His research interests include software-defined networking, cloud datacenter network, and performance optimization for networking algorithms and systems.

**Xiang Wang** (xiang-wang11@mails.tsinghua.edu.cn) received his PhD degree in 2015 from Department of Automation, Tsinghua University. He received his MS degree from the School of Software Engineering, University of Science and Technology of China in 2010 and BS degree from the School of Telecommunication Engineering, Xidian University, China in 2007. His research interests include software-defined networking, distributed system, and performance issues in computer networking and system architectures.

**Jun Li** (junl@tsinghua.edu.cn) received his PhD degree in Computer Science from New Jersey Institute of Technology (NJIT), USA, and MS and BS degrees in Control and Information from Department of Automation, Tsinghua University. He is currently a professor at Tsinghua University, and Executive Deputy Director of the Tsinghua National Laboratory for Information Science and Technology. Before rejoined Tsinghua University in 2003, he held executive positions at ServGate Technologies, which he co-founded in 1999. Prior to that, he was a senior software engineer at EX-AR and TeraLogic. In between of his MS and PhD studies, he was an assistant professor then lecturer in the Department of Automation, Tsinghua University. His current research interests mainly focus on networking and network security.

# A Software-Defined Approach to IoT Networking

Christian Jacquenet and Mohamed Boucadair

(France Telecom Orange, Cesson-Sévigné 35512, France)



## Abstract

It is foreseen that the Internet of Things (IoT) will comprise billions of connected devices, and this will make the provisioning and operation of some IoT connectivity services more challenging. Indeed, IoT services are very different from legacy Internet services because of their dimensioning figures and also because IoT services differ dramatically in terms of nature and constraints. For example, IoT services often rely on energy and CPU-constrained sensor technologies, regardless of whether the service is for home automation, smart building, e-health, or power or water metering on a regional or national scale. Also, some IoT services, such as dynamic monitoring of biometric data, manipulation of sensitive information, and privacy needs to be safeguarded whenever this information is forwarded over the underlying IoT network infrastructure. This paper discusses how software-defined networking (SDN) can facilitate the deployment and operation of some advanced IoT services regardless of their nature or scope. SDN introduces a high degree of automation in service delivery and operation—from dynamic IoT service parameter exposure and negotiation to resource allocation, service fulfillment, and assurance. This paper does not argue that all IoT services must adopt SDN. Rather, it is left to the discretion of operators to decide which IoT services can best leverage SDN capabilities. This paper only discusses managed IoT services, i.e., services that are operated by a service provider.



## Keywords

automation; dynamic service provisioning; Internet of Things; service function chaining; software-defined networking

## 1 Introduction

The Internet of Things (IoT) is a highly constrained, much larger networking infrastructure than legacy infrastructures that operators have known for decades. It is predicted there will be tens of billions of connected objects in the future, and IoT will be the de facto

networking infrastructure for a plethora of emerging services [1]. Some of these services are seen by many operators as key business development opportunities that need to be further explored or industrialized. Some IoT services are being deployed in the home and in dense urban environments. Other IoT services, such as e-health and energy distribution services, are being deployed on a regional, national or even interplanetary scale and require large-scale networking, computation, and storage.

IoT connectivity services rely on elementary functions such as forwarding and routing, quality of service (QoS), and security.

One of the main differences between IoT connectivity services and legacy connectivity services such as Internet access is the constrained nature of some of the technologies involved. For example, a wireless sensor network (WSN) deployed in an IEEE 802.15.4 [2] network environment assumes a maximum transmission unit (MTU) of 127 bytes, with only 80 bytes allocated to the MAC payload for an average 250 kbps rate.

A WSN includes sensors that are constrained in terms of CPU and energy. This can affect how IoT service-driven policies are designed and enforced, especially when the data being transported, e.g., personal biometric data, requires a high degree of privacy in the forwarding and routing schemes. In addition, IoT dimensioning figures suggest a very different, much larger networking scale. Several thousand connected devices, with or without route computation capabilities, are likely to be the norm rather than the exception in urban and regional areas and even nationwide (Table 1).

The design and operation of an IoT connectivity service is complicated by the inherent dynamics of the networking infrastructure. For example, connected devices may be rapidly (re)grafted onto or pruned from the IoT network infrastructure according to their CPU loads or remaining energy. They may also be (re)grafted onto or pruned from the IoT network infrastructure because they are in motion, e.g., biometric sensor bracelets [3], they have been damaged by weather, or they have entered sleep mode.

The deployment of a wide range of IoT services—from “smart home” residential services and automated building services to advanced personal e-health services—has become a

▼ Table 1. What makes IoT routing special

Internet Routing	IoT Routing
Nodes are routers	Nodes can be anything—sensors, actuators, routers, etc.
A few hundred nodes per network	1000+ nodes per network, depending on the nature of the service
Links and nodes are stable over time	Links are highly unstable and degrade communication. Nodes fail more often, e.g., exhausted batteries and CPU overload
No stringent routing constraints	Highly constrained environment
Routing is by default not application-aware	Routing must be application-aware, e.g., e-health services generate traffic that requires a high degree of privacy whereas energy-distribution services generate traffic that primarily requires low-latency routes

## A Software-Defined Approach to IoT Networking

Christian Jacquenet and Mohamed Boucadair

key strategy for operators. Such IoT services open up tremendous opportunities for operators to develop their businesses. The simultaneous development of cloud infrastructures and the introduction of automation techniques for service delivery and operation will likely boost IoT services.

Operators see IoT services as a key factor affecting business development and existing network infrastructures, from both a design standpoint and operational standpoint. The introduction of several hundred or even thousands of connected devices will distort the global routing system and affect traffic forwarding in access infrastructures but must not jeopardize the quality of legacy services.

Such effects are not only assessed from a dimensioning perspective, i.e., moving from several hundred network devices to several thousand connected objects with computing resources, but also from a traffic taxonomy perspective. IoT services typically demand the ability to compute (traffic-engineered) paths that can accommodate privacy characteristics of traffic. IoT services also involve other considerations that lead to complex, likely multimetric, multiconstrained routing objective functions that differ from current routing policies based on the classical hop-by-hop forwarding scheme.

Also, cloud-based resources, such as IoT service platforms, also affect the way IoT services are designed and operated. Operators now need to have skills in IT/network convergence, which suggests that current service delivery and operational procedures may need to be revised. The evolution of organizational practices is further affected by the introduction of advanced cross-platform, cross-segment residential, e-health, urban and corporate IoT services. These inevitably create new challenges because they have specific functional capabilities.

Software-defined networking (SDN) enables flexible, robust, scalable design and operation of IoT services. This paper discusses an original approach in which SDN is not limited to dynamic IoT resource allocation. IoT-specific policy provisioning information is exchanged between the SDN computation logic and some of the IoT service functions involved in the delivery and operation of the IoT service.

The proposed approach has a much broader scope: it can be used to dynamically expose and negotiate the parameters of an IoT service, and it can be used to assess whether the IoT services that have been dynamically delivered comply with what has been negotiated with the IoT application or service customer. This global, systemic, software-defined IoT networking approach is unique.

This paper is organized as follows. The following section introduces two cases where the design and operation of the IoT service are complicated by dimensioning and the nature of the traffic generated. Then, the paper discusses the benefits of SDN to IoT service delivery and operation. Furthermore, it discusses the nature of the various SDN building blocks used in the IoT service delivery procedure—from dynamic IoT service parameter exposure and negotiation to IoT resource allocation

and service fulfillment. The conclusion discusses what could be next for SDN-based IoT networking and what could be the role of network operators and service providers in this area.

## 2 Two Use Cases

Here we introduce two IoT services that are typically in the portfolio of a service provider. They are also prime examples of the complexity involved in smartly combining very different elementary capabilities, i.e., service functions that are usually supported by network elements. Besides basic forwarding capabilities, these services can usually manipulate privacy data, which often affects how connected devices dynamically compute and select routes to convey IoT traffic.

These cases create specific challenges in terms of scale but also in terms of QoS. Forwarding of biometric data collected by e-health sensors to the nearest hospital requires robust, low-latency routes whereas forwarding of power meter readings for billing purposes requires more reliable routes so that data does not need to be retransmitted.

The different routing objectives in the following two cases imply the need for an advanced, presumably multimetric route-computation logic that is not only fed specific service requirements and constraints but also proactively (or reactively) adapts to any event that may alter the network conditions in a deterministic, scalable manner. In this way, IoT services cannot be disrupted.

### 2.1 E-Health Services

A typical service that illustrates the challenges raised by IoT is e-health. In some contexts, e-health may require a network infrastructure that is highly reliable and preserves data integrity. Unlike some IoT services, where connected devices are only responsible for sending data, some e-health services may require traffic bi-directionality, perhaps for receiving check instructions and tweaking threshold settings.

In some e-health scenarios, monitoring a set of biometric data may involve dynamically computing routes for conveying data (collected by the sensors) to the nearest hospital when a threshold has been reached or selecting the hospital that can provide the most suitable specialist care. Given the sensitive nature of biometric data and the need to rapidly react to health emergencies, such as a heart attack, specific constraints should be overcome by the underlying forwarding and routing schemes.

These constraints can be overcome by dedicated traffic engineering, such as dynamic route computation, that takes into account not-so-usual routing metrics, such as the nature of the traffic, energy or CPU consumption of the communication device, or network bandwidth resources.

Also, there are typical seasonal epidemics, such as the winter flu, that need to be dynamically monitored on a regional or even national scale so that authorities can take appropriate ac-

tion (e.g., launch a vaccine campaign targeting people at risk).

Moreover, dynamic monitoring of an epidemic requires carefully designed traffic-forwarding policies adapted to manage mobile communities that process emergency calls and collect statistics on the importance, severity, and scope of the epidemic.

These two examples of e-health services create network challenges in terms of:

- reliable identification and efficient addressing and naming schemes for many connected devices (typically health sensors)
- dynamic, multimedric, self-adaptive route-computation schemes for service performance, scalability and robustness
- privacy preservation, so that sensitive data is not leaked to illegitimate nodes or data consumers
- dynamic mobility management and self-adaptive interconnected design schemes that leverage existing network infrastructure (both wired and wireless) for the sake of service-inferred traffic-forwarding policies.

Indeed, e-health services that dynamically monitor biometric data are available to users who may be mobile. As such, monitoring traffic-forwarding policies should be able to take advantage of available network infrastructures. Network interconnects may be needed to forward traffic upstream in the network or ensure that commands sent by an actuator connected somewhere on the Internet are reliably transmitted to the relevant connected devices. These network interconnects should be able to accommodate various kinds of IoT traffic envelopes and ensure such traffic can coexist with other types of traffic in order to minimize the risk of service disruption.

Self-adaptation can then be implemented according to the nature of the service to be delivered and the subsequent resource allocation decisions, e.g., route computation and bandwidth reservation.

## 2.2 Energy Management and Distribution

Dynamic management of energy distribution is another area where large-scale IoT might be used. Data collected from power meters is forwarded to metro agencies (perhaps for billing) but also contributes to the management of energy distribution during peak seasons, such as winter.

Forwarding the corresponding traffic requires capillary and WSNs that are connected with metropolitan and core networks (assuming both wired and wireless infrastructures).

Because of the nature of this traffic, adequate traffic engineering policies have to be enforced. This ensures that the computed paths will not only accommodate the type and amount of available resources but also the typical traffic patterns—e.g., N:1 or P:1 group communication schemes as a function of traffic directionality; sensor-collected data forwarded to metro, regional, or national energy control centers; or commands generated by an energy-control center and forwarded to a group of sensors so that energy consumption can be bet-

ter regulated.

This use case involves additional challenges besides those already mentioned for the deployment of robust e-health services. These challenges are related to:

- designing and dynamically enforcing multicast/broadcast traffic engineering policies on a large scale
- assessing the effect of corresponding traffic growth on the performance and scalability of core networking infrastructures from both a design and operation perspective. This results in the development of adapted traffic-forwarding paradigms.
- dynamically managing available bandwidth resources, such as radio channels in 802.15.4e environments.

## 3 Software-Defined Networking Can Help

The nature of some of IoT services encourages operators to be particularly flexible and agile during the service-delivery and operation phases. Some capabilities, such as firewall, that are needed to create, deliver, and maintain a feature of an IoT service may be hosted in various platforms typically located in a cloud infrastructure. Other capabilities, such as traffic forwarding and QoS, may be supported by in-network nodes such as dedicated service cards or devices with dedicated hardware.

Selection of capabilities needed to dynamically orchestrate and deliver an IoT service therefore benefits from the flexibility of cloud-hosted service platforms and applications coupled with SDN techniques [4] that include dynamic service-inferred IoT resource allocation and policy enforcement as well as feedback mechanisms for IoT service fulfillment and assurance.

In recent years, SDN-related activities have mostly centered on how a logically centralized SDN computation logic, often designated as an SDN controller or orchestrator, can provide network devices with configuration information pertaining to the various features required to deliver a (connectivity) service.

Also recently, the application of SDN to IoT networking has been investigated [5], [6]. However, the focus has primarily been on dynamically enforcing a traffic-forwarding policy within an IoT network infrastructure according to abstract models and virtualized functions.

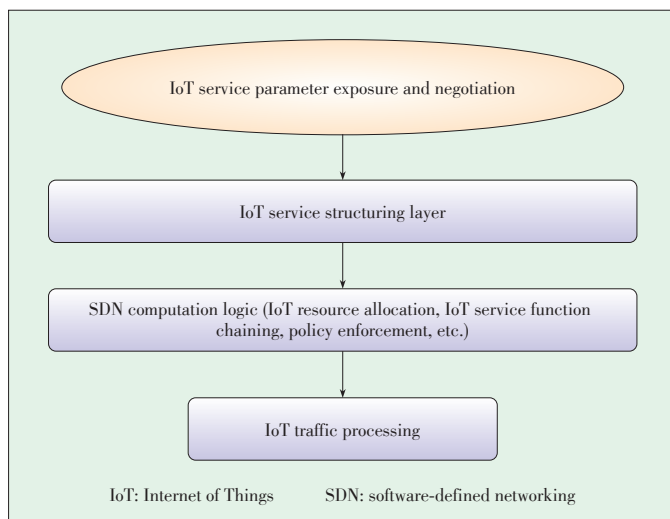
SDN combined with network function virtualization (NFV) and mass data analytics is a promising option for introducing high-degree automation into the overall IoT service-delivery procedure (**Fig. 1**)—from dynamic exposure and negotiation of IoT service parameters to resource allocation, policy enforcement, and service fulfillment and assurance.

Mass data analytics is required to optimize data aggregation and interpretation. SDN and data analytics can be used together to react to specific events and observed behaviors of the IoT underlying infrastructure. For example, they can be used to propose automated forwarding behaviors when there is an overload or failure. SDN and data analytics can also be used to offload some functions from the sensors and mediation servers



## A Software-Defined Approach to IoT Networking

Christian Jacquenet and Mohamed Boucadair



▲ Figure 1. IoT service delivery procedure.

while meeting real-time requirements of data processes required by some IoT services. Because time synchronization is critical for some data retrieval, SDN can be used to synchronize the clocks of involved nodes.

With NFV techniques, SDN can instantiate new IoT controllers and concentrators whenever required and wherever they are located in the transport network. In this way, data received from connected devices can be handled appropriately. The location and dimensioning of these controllers are automatically fed by SDN intelligence, which is based on various service-specific criteria that reflect the business guidelines of the IoT service provider.

An SDN platform can be used to manage one or more IoT services. Whether one or several SDN controllers are required in a given network depends on the deployment strategy, which has to take into account the number, nature, and scope of the IoT services to be delivered. Although the application of SDN techniques to IoT services is attractive, the approach discussed in this paper does not necessarily benefit each and every IoT service. Rather, we suggest that a software-defined approach to IoT networking is primarily beneficial for IoT services that require sophisticated treatment and processing.

Sensors are no longer application-dependent and can be customized for an application. SDN can significantly help customize involved nodes at large to accommodate the design requirements of an IoT service portfolio, from smart home automation to advanced e-health or energy distribution services.

Structurally, IoT services often rely on complex, multifunctional network architectures that involve on-field hardware with embedded software, connectivity distributed systems, cloud software components, and third-party developers. Related challenges include: constrained resources, occasional massive amounts of signaling information, queries, and reduced computational resources. A typical IoT network of several thousand nodes (Table 1) requires new data processing schemes,

stream processing, filtering, aggregation, and data mining.

## 4 IoT-Adapted SDN Mechanics

### 4.1 Dynamically Exposing and Negotiating IoT Service Parameters

An IoT connectivity service parameter (standard) template can be used for the dynamic negotiation procedure between a customer and IoT service provider [7]. In a biometric data-monitoring service that typically demands very low latency and privacy-preserving routes, such a template would include clauses about:

- sensor geolocation information, so that the SDN can find the most suitable routes to the nearest dispatch emergency center in a reliable and secure manner
- communication schemes and traffic patterns, e.g., a typical 1:N hose model where commands to collect biometric data during a daily duty cycle can be sent to  $N$  sensor bracelets from a controller in a monitoring center
- QoS guarantees and availability requirements, which may be expressed in terms of traffic loss or one-way delay metrics
- traffic isolation and privacy requirements. This typically requires encryption to ensure the privacy of personal data generated by biometric services.
- flow identification information, e.g., the IPv6 source address used by a given sensor to send data
- any relevant activation means (perhaps to dynamically graft a sensor to a specific Destination-Oriented Directed Acyclic Graph (DODAG) in a WSN that has Routing Protocol for Low Power and Lossy Networks (RPL) enabled [8].

### 4.2 Designing an IoT Service and Dynamically Allocating Resources

IoT resources can be dynamically selected and allocated according to the outcomes of the IoT service parameter negotiation and according to the information maintained by the SDN computation logic (Policy Decision Point) in an IoT resource repository, which stores the relevant IoT service data models [9].

Notifications originating from the IoT network may also affect the decision-making process of the SDN computation logic. For example, a sensor notifies the SDN computation logic that a 50% energy threshold has been reached, which leads to a decision to restrict it to only computing routes that are robust and reliable.

In the biometric data monitoring example mentioned previously, the outcomes of the service parameter negotiation feed the SDN computation logic, which derives the Objective Function [10] that locates the nearest grounded root in an RPL network environment. This grounded root could be hosted in a cloud service platform managed by the IoT service provider on behalf of the emergency dispatch center.

Depending on which RPL metrics best accommodate the IoT

service parameter negotiation results, the resulting DODAG topology may then look like either a high-quality link, battery-free routing environment, or low-latency link routing environment.

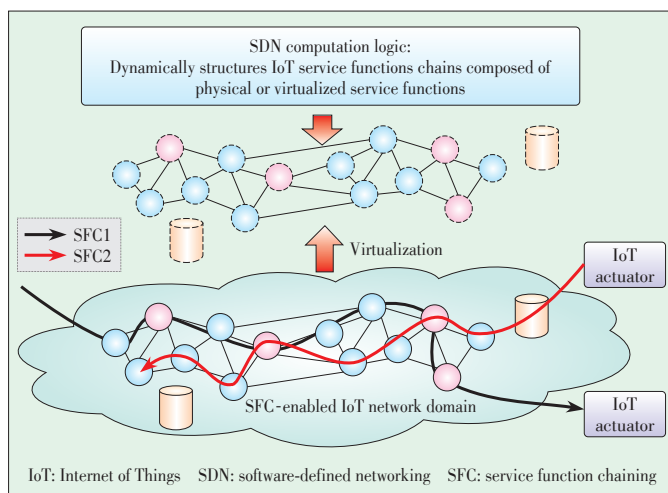
#### 4.3 Dynamically Structuring IoT Service Function Chains

To differentiate traffic handling in an IoT infrastructure, SDN-computed service function chaining techniques may be used [11]. These techniques are designed to enforce differentiated traffic-forwarding policies within the IoT network infrastructure and satisfy a set of service-specific IoT requirements, such as delegated encryption, security control, traffic shaping and scheduling, message formatting (add/remove field, versioning, protocol adjustment), or privacy preservation. In such contexts, the SDN computation logic dynamically structures the various service function chains according to service requirements that need to be satisfied to deliver a specific IoT service.

In the biometric data monitoring example, a set of elementary service functions need to be invoked. Such functions include sleep mode and sensor duty cycle management, to optimize energy consumption in particular; encapsulation and MTU management, to adapt to various network environments (especially when traffic needs to reach an IoT controller located upstream in the network); and security management, to preserve data privacy.

**Fig. 2** shows how two SDN-structured IoT service chains—SFC1 and SFC2—that are applied to traffic that crosses the IoT SFC domain.

- SFC1 = {Deep Packet Inspection (DPI), 6LoWPAN encapsulating capability [12], RPL DODAG Information Object (DIO) trickle timer and Destination Advertisement Object (DAO) route lifetime settings, TLS Proxy, 6Lo decapsulating capability}
- SFC2 = {DPI, 6Lo near field communication (NFC) encapsulating capability, expected transmission count (ETX) setting, auto ACK enforcement, CoAP/HTTP proxy, 6Lo decapsulating capability}



▲ Figure 2. SDN-computed IoT service function chaining [6].

ing capability}.

The IoT infrastructure is operated according to policies that tell IoT devices which flows are to be bound with which service chain.

#### 4.4 Dynamic Discovery of IoT Resources

An SDN approach involves a bootstrapping procedure for dynamic discovery of the IoT network topology (including active nodes), platforms, and their respective capabilities. This is necessary to feed the SDN computation logic.

The acquired information is stored and maintained in the resource repository. IoT service-driven policy provisioning and configuration information is derived from this repository and forwarded to the components that participate in the delivery and operation of an IoT service.

### 5 Virtualization Techniques Can Help Commoditize IoT Devices

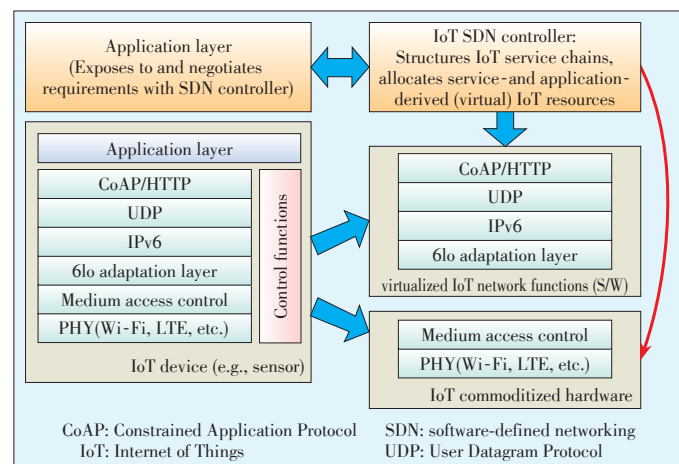
The lower layers, up to the Medium Access Control (MAC) layer, are embedded in commodity hardware. The upper layers, from the IPv6 network layer to the application layer (where Constrained Application Protocol (CoAP) [13] and HTTP reside) are virtualized and controlled by SDN (**Fig. 3**).

The SDN computation logic dynamically allocates virtual IPv6 forwarding and other RPL routing instances to master the flow of CoAP messages sent to a fleet of IoT devices for management purposes.

Such SDN-based deterministic flow mastery optimizes resource usage according to various parameters, such as location of IoT devices, whether these devices are mobile or not; acceptable ETX, to optimize duty cycle management; and data reception rate, to reduce energy consumption.

### 6 Conclusion and Next Steps

Combining SDN with virtualization is a likely precondition



▲ Figure 3. Virtualized IoT functions [6].

## A Software-Defined Approach to IoT Networking

Christian Jacquenet and Mohamed Boucadair

to the mass adoption of robust, scalable IoT services. IoT service-delivery and operational procedures can leverage SDN—from service parameter negotiation to resource allocation and invocation.

Alongside ongoing academic research on SDN in IoT networking, vendors and operators are developing IoT-adapted protocols and data models as well as the computation logic that lies beneath the SDN intelligence. These are areas where operators can contribute significantly in the years to come.

The SDN approach to IoT networking described in this paper is being further assessed through simulation and prototyping. The preliminary results of development activities on multi-metric IoT route computation, cross-platform IoT networking, and IoT-specific service function chaining will be communicated in 2016.

### References

- [1] O. Mazhelis, H. Warma, S. Leminen, *et al.*, "Internet-of-things market, value networks and business models: state-of-the-art report," University of Jyväskylä, Jyväskylä, Finland, Tech. Rep. TR-39, 2013.
- [2] J. T. Adams, "An introduction to IEEE STD 802.15.4," in *IEEE Aerospace Conference*, Big Sky, MT, USA, 2006. doi: 10.1109/AERO.2006.1655947.
- [3] N. Noury, A. Fleury, R. Nocua, *et al.*, "eHealth sensors, biomedical sensors, algorithms and sensor networks," *Innovation and Research in BioMedical Engineering*, IRBM vol. 30, no. 3, pp. 93–103, June 2009.
- [4] *Software-Defined Networking: A Perspective from within a Service Provider Environment*, IETF RFC 7149, Mar. 2014.
- [5] Z. Qin, G. Denker, C. Gianneli, *et al.*, "A software defined networking architecture for the internet-of-things," in *IEEE Network Operations and Management Symposium (NOMS)*, Krakow, Poland, 2014, pp. 1–9. doi: 10.1109/NOMS.2014.6838365.
- [6] M.-K. Shin, Y. Hong, and C. Y. Ahn, "A software-defined approach for end-to-end IoT networking," in *Proc. IETF91 SDRG Working Group Meeting*, Honolulu, USA, Nov. 2014.
- [7] *IP Connectivity Provisioning Profile (CPP)*, IETF RFC 7297, Jul. 2014.
- [8] *RPL: IPv6 Routing Protocol for Low Power and Lossy Networks*, IETF RFC 6550, Mar. 2012.
- [9] R. Sudhaakar and P. Zand, "6tiSCH resource management and interaction using CoAP," IETF, draft-ietf-6tisch-coap, Mar. 2015.
- [10] *Objective Function Zero for the Routing Protocol for Low-Power and Lossy Networks (RPL)*, IETF RFC 6552, Mar. 2012.
- [11] J. Halpern and C. Pignataro, "Service function chaining (SFC) architecture," IETF, draft-ietf-sfc-architecture, Aug. 2015.
- [12] IETF, *IPv6 over Networks of Resource Constrained Nodes (6lo) Working Group* [Online]. Available: <https://datatracker.ietf.org/wg/6lo/charter/>
- [13] *The Constrained Application Protocol (CoAP)*, RFC 7252, Jun. 2014.

Manuscript received: 2015-10-19

## Biographies

**Christian Jacquenet** (christian.jacquenet@orange.com) graduated from the Ecole Nationale Supérieure de Physique de Marseille, a French school of engineers. He joined Orange in 1989, and he is currently the director of the Strategic Program Office For Advanced IP Networking, Orange Labs. He is responsible for Orange's IPv6 program, which aims to define and drive the Group's IPv6 strategy. He also conducts development activities in the areas of software-defined networking and service function chaining. He has authored and co-authored several Internet drafts and IETF RFC standards on dynamic routing protocols and resource allocation techniques. He has also authored papers and books on IP multicasting, traffic engineering, and automated IP service delivery techniques.

**Mohamed Boucadair** (mohamed.boucadair@orange.com) is an IP networking strategist at France Telecom. He previously worked as a senior IP architect at FT and worked in the corporate division of FT, which made recommendations on the evolution of IP/MPLS core networks. He has worked for FT R&D and has been part of the team working on VoIP services. He has been involved in IST research projects, working on dynamic provisioning and inter-domain traffic engineering. He has also worked as an R&D engineer in charge of dynamic provisioning, QoS, multicast and intra/inter-domain traffic engineering. He has authored many journal articles and has written extensively on these subjects. He holds several patents on VoIP, IPv4 service continuity, and IPv6.

## Roundup

### Introduction to ZTE Communications

*ZTE Communications* is a quarterly, peer-reviewed international technical journal (ISSN 1673-5188 and CODEN ZCTOAK) sponsored by ZTE Corporation, a major international provider of telecommunications, enterprise and consumer technology solutions for the Mobile Internet. The journal publishes original academic papers and research findings on the whole range of communications topics, including communications and information system design, optical fiber and electro-optical engineering, microwave technology, radio wave propagation, antenna engineering, electromagnetics, signal and image processing, and power engineering. The journal is designed to be an integrated forum for university academics

and industry researchers from around the world. *ZTE Communications* was founded in 2003 and has a readership of 5500. The English version is distributed to universities, colleges, and research institutes in more than 140 countries.

It is listed in Inspec, Cambridge Scientific Abstracts (CSA), Index of Copernicus (IC), Ulrich's Periodicals Directory, Chinese Journal Fulltext Databases, Wanfang Data — Digital Periodicals, and China Science and Technology Journal Database. Each issue of *ZTE Communications* is based around a Special Topic, and past issues have attracted contributions from leading international experts in their fields.

# ***ZTE Communications Guidelines for Authors***

## **• Remit of Journal**

*ZTE Communications* publishes original theoretical papers, research findings, and surveys on a broad range of communications topics, including communications and information system design, optical fiber and electro-optical engineering, microwave technology, radio wave propagation, antenna engineering, electromagnetics, signal and image processing, and power engineering. The journal is designed to be an integrated forum for university academics and industry researchers from around the world.

## **• Manuscript Preparation**

Manuscripts must be typed in English and submitted electronically in MS Word (or compatible) format. The word length is approximately 3000 to 8000, and no more than 8 figures or tables should be included. Authors are requested to submit mathematical material and graphics in an editable format.

## **• Abstract and Keywords**

Each manuscript must include an abstract of approximately 150 words written as a single paragraph. The abstract should not include mathematics or references and should not be repeated verbatim in the introduction. The abstract should be a self-contained overview of the aims, methods, experimental results, and significance of research outlined in the paper. Five carefully chosen keywords must be provided with the abstract.

## **• References**

Manuscripts must be referenced at a level that conforms to international academic standards. All references must be numbered sequentially in-text and listed in corresponding order at the end of the paper. References that are not cited in-text should not be included in the reference list. References must be complete and formatted according to *ZTE Communications* Editorial Style. A minimum of 10 references should be provided. Footnotes should be avoided or kept to a minimum.

## **• Copyright and Declaration**

Authors are responsible for obtaining permission to reproduce any material for which they do not hold copyright. Permission to reproduce any part of this publication for commercial use must be obtained in advance from the editorial office of *ZTE Communications*. Authors agree that a) the manuscript is a product of research conducted by themselves and the stated co-authors, b) the manuscript has not been published elsewhere in its submitted form, c) the manuscript is not currently being considered for publication elsewhere. If the paper is an adaptation of a speech or presentation, acknowledgement of this is required within the paper. The number of co-authors should not exceed five.

## **• Content and Structure**

*ZTE Communications* seeks to publish original content that may build on existing literature in any field of communications. Authors should not dedicate a disproportionate amount of a paper to fundamental background, historical overviews, or chronologies that may be sufficiently dealt with by references. Authors are also requested to avoid the overuse of bullet points when structuring papers. The conclusion should include a commentary on the significance/future implications of the research as well as an overview of the material presented.

## **• Peer Review and Editing**

All manuscripts will be subject to a two-stage anonymous peer review as well as copyediting, and formatting. Authors may be asked to revise parts of a manuscript prior to publication.

## **• Biographical Information**

All authors are requested to provide a brief biography (approx. 100 words) that includes email address, educational background, career experience, research interests, awards, and publications.

## **• Acknowledgements and Funding**

A manuscript based on funded research must clearly state the program name, funding body, and grant number. Individuals who contributed to the manuscript should be acknowledged in a brief statement.

## **• Address for Submission**

magazine@zte.com.cn  
12F Kaixuan Building, 329 Jinzhai Rd, Hefei 230061, P. R. China



# **ZTE COMMUNICATIONS**



**ZTE Communications has been indexed in the following databases:**

- Cambridge Scientific Abstracts (CSA)
- China Science and Technology Journal Database
- Chinese Journal Fulltext Databases
- Inspec
- Ulrich's Periodicals Directory
- Wanfang Data—Digital Periodicals