Real-time UHD Scalable multi-layer HEVC encoder architecture

Ronan Parois
ATEME, Paris (France)
Email: r.parois@ateme.com

Wassim Hamidouche
IETR INSA Rennes, (France)
Email: whamidou@insa-rennes.fr

Elie Gabriel Mora and Mickael Raulet
ATEME, Paris (France)
Email: e.mora@ateme.com
m.raulet@ateme.com

Olivier Deforges
IETR INSA Rennes, (France)
Email: odeforge@insa-rennes.fr

***Abstract*—The High Efficiency Video Coding (HEVC) standard enables meeting new video quality demands such as Ultra High Definition (UHD). Its scalable extension (SHVC) allows encoding simultaneously different versions of a video, organised in layers. Thanks to inter-layer predictions, SHVC provides bit-rate savings over an equivalent HEVC simulcast encoding. Therefore, SHVC seems a promising solution for both broadcast and storage purposes.**

This paper proposes a multi-layer architecture of a pipeline of software HEVC encoder to perform real-time UHD spatially-scalable SHVC encoding. Inter-layer predictions are furthermore implemented to provide bit-rate savings with a minimum impact on complexity. The proposed architecture provides a good trade-off between coding gains and coding speed achieving real-time performance for 1080p60 and 1600p30 sequences in 2× spatial scalability. Moreover, experimental results show more than a 1000× speed-up compared to the SHVC reference software (SHM) and an introduced delay only reaching 14% of the equivalent HEVC coding speed.

I. Introduction

Nowadays, video coding standards have to face the increasing demands for new video formats. The newly developed High Efficiency Video Coding (HEVC) [1] standard allows, in particular, the rapid deployment of emerging services in UHD [2]. The HEVC standard was defined by the Joint Collaborative Team on Video Coding (JCT-VC) in early 2013. It is designed and particularly adapted to the encoding of high spatial resolution video. The HEVC frame subdivision into variable length Coding Tree Units (CTU) and Coding Units (CU) enables significant coding efficiency improvements [3] compared to its predecessors MPEG-2, H.263 and H.264/Advanced Video Coding (AVC) [4]. HEVC is also the first video standard that provides specific parallelism tools to improve the coding speed including tiles and wavefront concepts [5].

Several HEVC extensions [6] have been defined to support additional coding features such as the spatial, bit-depth, quality, color gamut and codec scalability with the Scalable High efficiency Video Coding (SHVC) extension [7]. Previous standards also defined a scalable extension like Scalable Video Coding (SVC) [8] for the H.264/AVC with already known performance [9]–[11]. As for HEVC with the HEVC reference software (HM), a Scalable HEVC reference software (SHM) provides an encoding and decoding solution for SHVC. Comparisons of these softwares have been presented in several publications [12]–[14] with a High Definition (HD) limitation for SHVC. Since those experiments show that SHVC is a promising solution for future video broadcasting and storage purposes, there is a need for a real-time SHVC encoder.

In this paper, we propose a spatially-scalable software SHVC encoder performing real-time encoding of two layers on 1080p60 and 1600p30 sequences from the Common Test Conditions (CTC) [15]. This solution is based on a real-time HEVC encoder developed by ATEME, a French company known for its expertise in video coding and its presence in contribution and distribution markets. To the best of our knowledge, the proposed encoder is the first real-time and parallel spatially-scalable SHVC encoder for HD and 2K video sequences at different bit-rates. There are real time and parallel software SHVC decoders [16], [17] but, to our knowledge, there are no SHVC encoder except SHM.

The rest of this paper is organized as follows. The SHVC extension and its applications are presented in Section II. The architecture of the proposed real time software SHVC encoder is detailed in Section III. Section IV presents and analyses the performance of the proposed real-time and parallel SHVC encoder. Finally, Section V concludes this paper.

II. SHVC extension

Scalable video coding consists in coding a video in several layers, each layer corresponds to a different quality level of this video. These quality levels may represent different spatial or temporal resolutions, quality, bit depths, color gamuts, or even video coding standards. The first layer, called Base Layer, represents the lowest quality level to be encoded. The Enhancement Layers (ELs) represent the upper quality level and use coding information from lower layers (BL or lower ELs). The video layers are multiplexed in a single scalable video bitstream. In this paper, we focus on two layers spatial scalability in 2× ratio.

A. Scalability in SHVC

The SHVC encoder can practically be structured in two ways. The first structure consists in a single encoder where all layers are processed at the same time. The second one consists

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

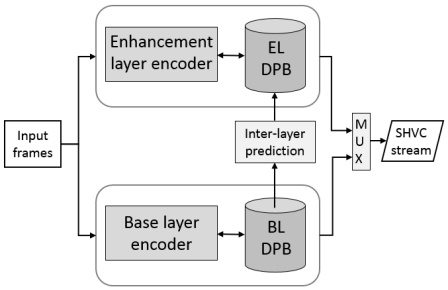

Fig. 1. SHVC encoder in 2-layers structure

in instantiating as many single layer encoders as desired and interface them to perform inter-layer predictions.

Whatever the employed structure, each layer processes the same video sequence in a different spatial resolution. Once the coding information of every layer is available, all this information is multiplexed and interlaced in a single SHVC bitstream. This operation is performed using high-level syntax and is detailed in [7]. This coding information may also be used for inter-layer prediction. Figure 1 represents the SHVC architecture with two layers. As illustrated, at least two interface operations are required between single layers: a multiplexing operation and an inter-layer prediction. While multiplexing only consists in interlacing syntax elements, inter-layer prediction is computationally more complex.

B. Inter-layer prediction

In SHVC extensions of video standards, coding efficiency in upper layers is achieved using coding information from lower layers. BL can only be used as a reference layer while ELs are predicted layers and may also be reference layers. This process, called inter-layer prediction, is usable in addition to the conventional intra-layer predictions modes (inter and intra). More precisely, inter-layer prediction allows using information (the reconstructed texture, the motion information and/or the residual) of the corresponding Prediction Unit (PU) from a lower layer to encode the current PU in a predicted layer. In the case of spatial scalability, the block used as reference for inter-layer predictions is up-sampled to the predicted layer size equivalent through an up-sampling operation. This operation use 8-tap filter and 4-tap filter for luma and chroma picture components, respectively. Therefore, the complexity of the SHVC encoder, compared to a single layer HEVC encoder, is increased by both the encoding of the BL as well as the up-sampling operations [17].
When using a reconstructed lower-layer texture, the reconstructed picture is copied from the Decoded Picture Buffer (DPB) of the lower-layer to the predicted layer DPB and is signaled as a long term reference when inter-layer syntax elements are activated. A specific reference index is used for signaling inter-layer reference pictures.

C. Applications

Since the video in scalable extensions of video standards are coded at different quality levels, scalable video coding allows adapting the video bitstream according to network conditions and decoder capabilities. Applications with such needs are mainly video conferencing and video streaming for instance [18]. Moreover, scalable video coding provides a better error resiliency. When network errors lead to data loss in the EL, video quality is less impacted than what it would have been on an equivalent single layer.

SHVC alternatives include simulcast and transcoding approaches. Transcoding consists in decoding a video bitstream before re-encoding it with other encoding parameters. Compared to the scalable coding, transcoding adds significant delay, bandwidth and complexity overheads. Simulcast consists in encoding the same video at different quality levels, using a different single-layer encoder for each layer. The encodings are completely independent (no inter-layer prediction). Selected simulcast levels are chosen by decoders depending on their capabilities and/or network conditions.

III. Proposed SHVC encoder

In this section, we present the initial single layer real-time HEVC encoder used as our SHVC basis. Then we describe the proposed structure for our SHVC software encoder.

A. HEVC basis encoder

Our proposed SHVC solution is based on a multi-support real-time software encoder developed by ATEME that we refer to as single layer encoder ATEME (SLE-ATEME) in the rest of this paper. SLE-ATEME is able to perform different standard encodings from MPEG-2 to H.265/HEVC. Real-time is guaranteed thanks to various Single Instruction Multiple Data (SIMD) optimizations and, for more recent standards, parallelism. While used as an HEVC encoder, SLE-ATME parallelism is mostly performed using tiles [19] with only one slice per frame. It also integrates a bit-rate control module which allows to perform encodings at constant quantizer and Constant Bit-Rate (CBR).

An SLE-ATEME encoding is based on a pipeline process applied at the frame level. This pipeline is illustrated in Figure 2 where steps preceding the coding decision correspond to various operation such as frame acquirement (from a file or from an acquisition device), various pre-processing operations, Group Of Picture (GOP) construction or motion estimation. At the end of the pipeline, the reconstructed frame ready for inter prediction is available for encoding the next coded pictures. This encoding process introduces a latency that depends on the type of the frame to be encoded. For example, while encoding 1920×1080 8-bit video, a full pipeline (i.e latency) lasts 15 frames. All pipeline steps have almost the same complexity except the last one: the coding decision, which is usually more complex. This decision relates to quadtree structure, prediction modes, prediction information and transform information. At the end of the coding decision, all coding information is available including the reconstructed frame usable for inter-prediction. Once the first reconstructed frame is available, it takes only one pipeline step to reconstruct the following

frame in coding order. As a result the coding decision duration defines the encoding cadence (i.e. encoding rate).

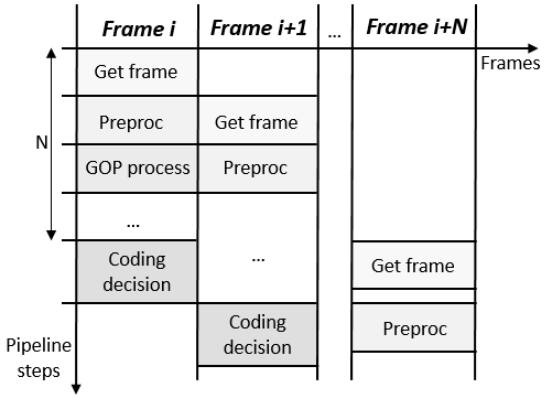

Fig. 2. SLE-ATEME pipeline

B. Proposed SHVC structure

The proposed SHVC encoder, which we refer to in the rest of this paper as SHVC-ATEME, is composed of two SLE-ATEME. Each SLE-ATEME processes a different resolution for a spatial scalability application with two layers: a BL and an EL. Since layers are processed simultaneously, a synchronization between the two encoders is required. This synchronization is performed at the pipeline level. Once each pipeline step is processed, a synchronization signal is sent to the management thread. The next pipeline step starts once all signals are received. As a result, the cadence is defined by the most complex step.

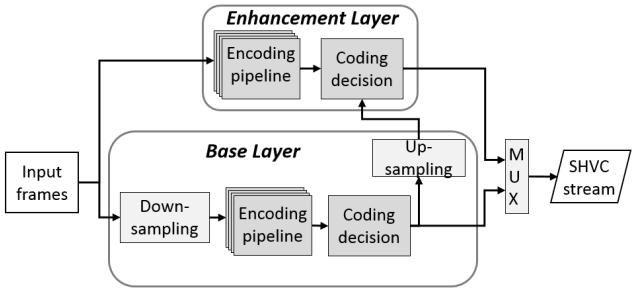

Fig. 3. SHVC-ATEME architecture

The SHVC-ATEME encoder has only one input, meaning that a down-sampling filter is required before encoding lower layer. This down-sampling operation is performed in the BL encoder at the very beginning of the pipeline as illustrated in Figure 3. This filter adds a pipeline step to the BL while the EL pipeline remains unchanged.

An up-sampling filter is also added to the BL pipeline for inter-layer prediction. This filtering is done once the reconstructed frame is available on the BL, at the end of the coding decision. The up-sampling operation is SIMD-optimized to minimise the additional delay introduced by this additional operation. In [17], the 2× up-sampling filter has been estimated to 17% of the total decoding time.

For inter-layer prediction purposes, the EL needs the up-sampled BL coding information before starting the coding decision. In the SHVC-ATEME encoder, we use the pipeline structure to fulfil this requirement: the EL coding decision should start a pipeline step after the BL coding decision.

Since the BL pipeline is longer by one step, a two-pipeline-step gap is introduced between BL and EL when encoding starts. Figure 4 illustrates this pipeline alignment requirement. With such a structure, only one pipeline step is added to the BL latency and only two to the EL. As a result, the SHVC-ATEME encoder latency and cadence are respectively only slightly higher and lower than the latency and cadence of the SLE-ATEME.

The SHVC-ATEME encoder performs inter-layer prediction with only I-slice references. Two main reasons explain this choice which is based on a trade-off between bit-rate savings and complexity. First, an HEVC I-slice predicted by an inter-layer reference becomes an SHVC P-slice. It is known that I-slices introduce a significant bit-rate cost compared to other slice types [1]. As a result, bit-rate saving would be more significant by adding an inter-layer reference to I-slices than to other slice types. Secondly, adding an inter-layer reference to other slice types would significantly reduce coding speed.

In SHVC-ATEME, coding speed performance relies on two parallelism levels. The first one is a frame-level parallelism where tiles are used to improve coding speed. This parallelism is directly inherited from SLE-ATEME. The second one is a layer-level parallelism since the BL encoder and the EL encoder are processed simultaneously. This parallelism is the consequence of the choice we made on the employed structure: an encoder per layer. Moreover, this structure allows to easily increase the number of layers with minimal impact on the global SHVC encoding complexity.

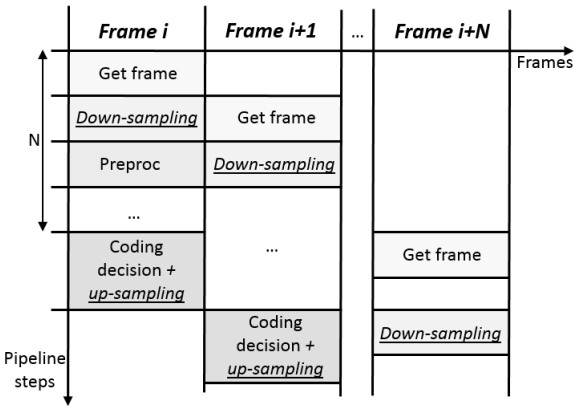

Fig. 4. SHVC-ATEME pipeline

IV. Results

A. Experimental settings

The platform used for experiments is composed of four Intel Xeon E5-4627V3 CPUs. It provides 4x8 cores (32 threads)

TABLE I
COUPLE QUANTIZER USED FOR TESTING

BL	22	22	26	26	30	30	34	34
EL	22	24	26	28	30	32	34	36

TABLE II
TESTS SEQUENCES

Seq.	Seq. name	Resolution	Frame rate
A1	Traffic	2560×1600	30
A2	PeopleOnStreet	2560×1600	30
B1	Kimono	1920×1080	24
B2	ParkScene	1920×1080	24
B3	Cactus	1920×1080	50
B4	BasketballDrive	1920×1080	50
B5	BQTerrace	1920×1080	60

each running at 3.30 GHz.

The sequences used in for experiments are presented in Table II. We consider class A and class B sequences defined in the Common Test Conditions (CTC) [15] which contain 1600p sequences as the highest resolution. The used SHVC encoders are SHVC-ATEME and SHM version 7.0 [20] in 2× spatial scalability configuration. For single layer equivalents, the HEVC encoders used are SLE-ATEME and HM version 14.0 [21]. The 2× spatial scalability configuration means that encoding a 1920×1080 (or 2560×1600) resolution format in HEVC corresponds to a 960×540 (1280×800) resolution format on the BL and a 1920×1080 (2560×1600) resolution format on the EL, respectively.

The experiments are conducted in a constant quantizer, all intra configuration defined in the CTC. Table I summarizes the pairs of quantizer used for BL and EL. In an all intra configuration, each layer only has one slice type to encode: I-slice for BL and P-slice for EL.

The performance of the proposed solution is assessed in terms of coding gains, measured with the Bjontegaard metric [22] and encoding frame rates for speed performance evaluation. When only the SHVC EL is compared to the single layer equivalent, coding gains correspond to the bit-rate saving brought by SHVC over HEVC at the same video resolution and quality.

TABLE III
CODING GAINS OF SHM AND SHVC-ATEME OVER HM AND SLE-ATEME RESPECTIVELY ON EL

Seq.	SHM vs HM	SHVC-ATEME vs SLE-ATEME
A1	-40.2%	-19.9%
A2	-44.1%	-23.1%
B1	-48.7%	-31.1%
B2	-34.2%	-19.3%
B3	-31.1%	-18.9%
B4	-25.7%	-17.3%
B5	-17.6%	-9.8%
Avg.	**-34.5%**	**-19.9%**

TABLE IV
SHVC ENCODERS CODING SPEED COMPARISON

Seq.	QP BL EL	Coding speed (fps) SHVC-ATEME	SHM
A1	22 22	39.8	22.4E-3
	34 36	49.3	31.6E-3
	avg. all QP	45.2	27.2E-3
A2	22 22	39.7	22.4E-3
	34 36	49.0	30.8E-3
	avg. all QP	45.2	26.8E-3
B1	22 22	78.9	60.0E-3
	34 36	96.0	81.5E-3
	avg. all QP	89.6	68.1E-3
B2	22 22	70.1	28.4E-3
	34 36	92.4	72.8E-3
	avg. all QP	82.9	53.4E-3
B3	22 22	70.4	42.5E-3
	34 36	94.2	72.4E-3
	avg. all QP	84.2	60.2E-3
B4	22 22	72.8	46.7E-3
	34 36	96.6	62.3E-3
	avg. all QP	87.0	59.1E-3
B5	22 22	65.9	43.2E-3
	34 36	92.4	71.7E-3
	avg. all QP	81.4	57.7E-3

TABLE V
AVERAGE SHVC-ATEME SPEED UP OVER SHM AND SLE-ATEME

Seq.	Speed-up over SHM	SLE-ATEME
A1	1667.9	0.87
A2	1694.4	0.86
B1	1384.6	0.87
B2	1690.0	0.87
B3	1414.8	0.88
B4	1482.5	0.87
B5	1423.6	0.88

B. Result analysis

Table III presents the bit-rate savings of SHM and SHVC-ATEME over HM and SLE-ATEME respectively. We can notice that spatial scalability brings bit-rate savings in all video tests sequences.

SHVC-ATEME provides less bit-rate savings than SHM. Figure 5 presents a rate distortion curve for HM, SHM,

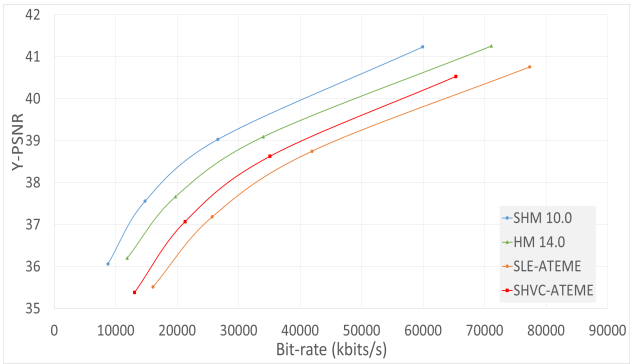

Fig. 5. HM, SHM and SHVC-ATEME Y-PSNR vs Bit-rate (B4 sequence)

SHVC-ATEME and SLE-ATEME on the B4 sequence with equal QPs on both layers. We can observe a significant difference between the SHVC-ATEME and the SHM bitrates, i.e. SHVC-ATEME requires a higher bit-rate than SHM to encode the same video at an equivalent quality. This important bit-rate difference partially explains the coding gain difference observed on Table III. Indeed, SLE-ATEME is not designed to encode all intra sequences (i.e it is less efficient than HM). The first target of SLE-ATEME is broadcast where an all intra configuration is not commonly used. As a result, in SLE-ATEME, some simplified coding decisions are made to achieve real-time performance at the price of a loss in coding efficiency. This is not the case of HM which always seeks out the most coding efficient decision.

Table IV shows the performance of the SHVC encoders in terms of encoding frame rate. We can observe that the proposed solution is able to encode all sequences in real-time. Table V presents the SHVC-ATEME speed-up over SHM and SLE-ATEME. The SHVC-ATEME encoder is able to offer more than a 1000× speed-up factor for each encoded sequence. As expected, a delay is introduced between SHVC-ATEME and SLE-ATEME, but it only represents 14% of the SLE-ATEME cadence.

Moreover, the SHVC-ATEME encoder has been tested in a live setup. The SHVC encoder is interfaced with an acquisition device supporting up to 4Kp60 resolution [23]. This setup introduces no delay in frame acquisition. As a result, we were able to test real-time encoding of 4Kp30 sequences (not provided in the CTC). This experiment has been performed at ATSC 3.0 demonstrator [24].

V. Conclusion

In this paper, we have presented an SHVC encoder implementing inter-layer predictions to achieve coding efficiency. Inter-layer predictions are only used on I-slice references to balance bit-rate savings and coding speed. Moreover, with our single layer pipeline HEVC basis, we introduce very little latency with more encoded layers. Experiments show that the proposed SHVC solution is able to perform real-time encodings in 1600p resolution from the CTC. The solution also demonstrated real-time performances on UHD sequences random access configuration at the ATSC 3.0 meeting [24] for which SHVC is a challenging video coding solution candidate [25].

In future, Author will focus on random access encoding configuration to get closer to the nowadays broadcast environment.

References

[1] G. Sullivan, J. Ohm, W. 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, december 2012.

[2] ITU-R, "Recommendation ITU-R BT 2020, Parameter values for ultra-high definition television systems for production and international programme exchange," august 2012.

[3] 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, december 2012.

[4] T. Wiegand, G. 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. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1218189

[5] C. C. Chi, M. Álvarez-Mesaa, B. Juurlink, G. Clare, F. Henry, S. Pateux, and S. Thomas, "Parallel Scalability and Efficiency of HEVC Parallelization Approaches," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1827–1838, december 2012.

[6] G. J. Sullivan, J. M. Boyce, Y. Chen, J.-R. Ohm, C. A. Segall, and A. Vetro, "Standardized Extensions of High Efficiency Video Coding (HEVC)," *IEEE journal of selected topics in signal processing*, vol. 7, no. 6, pp. 1001–1016, december 2013.

[7] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC : Scalable Extensions of the High Efficiency Video Coding Standard," *IEEE Transaction on Circuits and Systems for Video Technology*, pp. 20–34, july 2015.

[8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for video Technology*, pp. 1103–1120, september 2007.

[9] S. Sanz-Rodrígueza, M. Álvarez-Mesaa, T. Mayerb, and T. Schierl, "A parallel H.264/SVC encoder for high definition video conferencing," *Signal Processing: Image Communication*, vol. 30, pp. 89–106, january 2015.

[10] H. Schwarz and T. Wiegand, "R-D Optimized Multi-Layer Encoder Control for SVC," *IEEE International Conference on Image Processing*, pp. II281–II284, october 2007.

[11] S. Hahm, C. Park, K. Park, and M. Kim, "Implementation on a real-time SVC encoder for mobile broadcasting," *5th IEEE Consumer Communications and Networking Conference*, pp. 1236–1237, january 2008.

[12] U. S. M. Dayananda and V. Swaminathan, "Investigating Scalable High Efficiency Video Coding for HTTP streaming," *IEEE International Conference on Multimedia and Expo Workshops*, pp. 1–6, june-july 2015.

[13] A. Kessentini, T. Damak, M. A. B. Ayed, and N. Masmoudi, "Scalable high efficiency video coding (SHEVC) performance evaluation," *World Congress on Information Technology and Computer Applications Congress*, pp. 1–4, june 2015.

[14] U.-K. Park, H. Choi, J. W. Kang, and J.-G. Kim, "Scalable video coding with large block for UHD video," *IEEE Transactions on Consumer Electronics*, pp. 932–940, august 2012.

[15] V. Seregin and Y. He, "Common SHM test conditions and software reference configurations," *Joint Collaborative Team on Video Coding, JCTVC-Q1009, Valencia, ES*, march - april 2014.

[16] W. Hamidouche, M. Raulet, and O. Deforges, "Parallel SHVC decoder: Implementation and analysis," *IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, july 2014.

[17] ——, "4K Real-Time and Parallel Software Video Dcoder for Multilayer HEVC Extensions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 169–180, january 2016.

[18] Y. Ye, Y. He, and X. Xiaoyu, "Manipulating Ultra-High Definition Video Traffic," *IEEE Multimedia*, pp. 73–81, july - september 2015.

[19] K. Misra, A. Segall, M. Horowitz, S. Xu, A. Fuldseth, and M. Zhou, "An Overview of Tiles in HEVC," *IEEE journal of selected topics in signal processing*, pp. 969–977, december 2013.

[20] "SHM reference software version 10.0," *https://hevc.hhi.fraunhofer.de*.

[21] "HM reference software version 14.0," *https://hevc.hhi.fraunhofer.de*.

[22] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," *ITU-T video Coding Experts Group document VCEG-M33*, march 2001.

[23] Dektec Digital Video BV, "DTA-2174 Leaftlet," June 2015.

[24] D. McAdams, "Sinclair Demos HDR 4KTV over ATSC 3.0 in Vegas," *http://www.tvtechnology.com/news/0002/sinclair-demos-hdr-4ktv-over-atsc-30-in-vegas/277546*, december 2015.

[25] R. Chernock and J. C. Whitaker, "Next-Generation Broadcast Television: ATSC 3.0," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 158–162, january 2016.