Efficient Parallel Architecture for a Real-time UHD Scalable HEVC Encoder

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

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

Jérôme Vieron
and Mickaël Raulet
ATEME, Paris (France)
Email: j.vieron@ateme.com
m.raulet@ateme.com

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

Abstract—The scalable extension (SHVC) of the High Efficiency Video Coding (HEVC) allows encoding in layers a video with multiple quality level such as resolution, bit-depth or Signal to Noise Ratio (SNR). Compared to the equivalent HEVC simulcast, the SHVC extension provides inter-layer prediction mechanisms enabling significant bit-rate savings. Moreover these inter-layer prediction mechanisms are less complex than those from former standards. Therefore, SHVC seems a promising solution for both broadcast and storage applications and is considered in the ATSC 3.0 as video coding solution. Indeed the spatial scalability is an application use-case considered in the ATSC 3.0.

This paper proposes a scalable multi-layer architecture combining pipelined software HEVC encoders. The proposed architecture provides a good trade-off between parallelism over layer and latency. Moreover two configurations are proposed for Live or File encodings with real-time or best coding efficiency targets, respectively. Results present a 2× spatial scalability application of this architecture achieving in a low-delay configuration real-time video encodings of 1080p60 and 1600p30 sequences. Moreover the proposed SHVC solution also demonstrated real-time encodings of UHD contents at an ATSC 3.0 meeting in random-access configuration.

I. INTRODUCTION

In some years the video consumption has exploded to a point where more and more new devices require new video usages. These new usages introduce more and more constraints that have to be faced by new video coding standards. The High Efficiency Video Coding (HEVC) [1] standard, defined by the Joint Collaborative Team on Video Coding (JCT-VC) in early 2013, enables the rapid deployment of emerging services of these new usages, such as UHD [2] for example. HEVC is designed and particularly adapted to the encoding of high spatial resolution video. As its predecessors MPEG-2, H.263 and H.264/Advanced Video Coding (AVC) [3], the HEVC standard uses a frame subdivision into blocks of pixel. However, the HEVC standard proposes another definition of this subdivision called Coding Tree Units (CTU). The CTUs are squared region with variable size including 64×64 and 32×32 and 16×16 pixels. These CTUs can also be divided into Coding Units (CU) used in the coding decision. Compared to former standards, the new subdivision enables significant improvements in terms of coding efficiency [4]. The HEVC standard also defines parallelism tools including tiles and wavefront parallel processing to enhance the coding rate [5].

Moreover HEVC allow repartition over groups of CTU in slices. Three slice types exist: I and P and B slices. I-slices only allow intra prediction with a spatial correlation. P and B-slices allow both inter and intra prediction for temporal and spatial correlation with respectively one and two list of references. A first HEVC extension [6] has been released in october 2014 enabling scalable features applied to video quality level. This scalable extension is called the Scalable High efficiency Video Coding (SHVC) [7] and enables scalable video coding including spatial and bit-depth and Signal to Noise Ratio (SNR) and color gamut and codec scalability features. The HEVC reference software (HM) [8] and the Scalable HEVC reference software (SHM) [9] provides both an encoding and decoding solution for HEVC and SHVC, respectively.

The different video quality levels enable the scalable video coding to adapt the encoded bitstream according to decoder capabilities and network conditions. For example, it enables preventing errors resilience from traffic congestions. These requirements can be found in applications such as video streaming and video conferencing applications for instance [10]. Alternatives to scalable video coding in such applications includes simulcast approach. Simulcast consists in encoding the same video at different video quality level. Each quality level corresponds to a layer and is independent from the other. For an equivalent fidelity, the simulcast encoding bit-rate is higher than with scalable video coding. As a result, for HD and UHD video contents, SHVC bit-rate savings over simulcast can reach from 16.5% to 29.6% and from 27.1% to 32.18% [7], respectively.

Previous experiments enhanced the SHVC coding efficiency with a normative improved scalable coding mode [11] and a non-normative adaptive rate-control solution [12]. Therefore, these previous works illustrate the SHVC promising potential as video broadcasting and storage solution. The fact that these works exist also show a need for SHVC encoder solution combining both real-time processing and coding efficiency.

Moreover the Advanced Television Systems Committee (ATSC) 3.0 [13] considers the SHVC as a video coding solution candidate. The ATSC defines a broadcast solution mainly employed in the US. This broadcast solution enables applications such as video conferencing that consider a real-time constraint on the employed video coding solution.

In previous works [14], [15] authors presented the employed architecture for scalable video coding with a spatial scalable application. In this paper, authors propose an enhancement of this spatially-scalable software SHVC encoder based on a professional real-time HEVC encoder. With an ATSC broadcast context, the proposed SHVC encoder performs real-time encodings. It integrates enhancements in the coding rate thanks to various optimisations on filtering operations that enable inter-layer prediction on I and P slices in real-time. To the best of our knowledge, there are real-time and parallel software SHVC decoders [16] but there is no real-time SHVC encoder. The proposed encoder is the first real-time and parallel spatially-scalable SHVC encoder addressing broadcast target.

The rest of this paper is organised as follows. The SHVC extension is presented in Section II. Section III details the proposed real-time SHVC encoder architecture and its optimisations. Performances in terms of coding rate and bit-rate savings of the solution are presented in Section IV. Finally, Section V concludes this paper.

II. SHVC EXTENSION

Scalable video coding consists in coding a video in several layers, each layer providing a specific quality level of this video. The quality level that is considered in our application is the video resolution for a spatial scalability. The lowest resolution is processed by the Base Layer (BL). The BL bitstream is HEVC compliant. The upper resolution are processed by Enhancement Layers (ELs). These ELs can used coding information from the BL or from lower ELs in a process called inter-layer prediction. Finally the coding information from each layers are combined to form a single multiplexed bitstream.

In SHVC, the layers encodings are mainly based on the HEVC standard but also require inter-layer operations: an inter-layer prediction and a multiplexing operation. The multiplexing operation enables the bitstream to be SHVC compliant by using high-level syntax. As detailed in [7], it consists only in interlacing NALUs (Network Abstraction Layer Units) from each layer. One the other hand, the inter-layer prediction is much more complex.

A. Inter-layer prediction process

In the SHVC standards, the prediction is based on the HEVC standards (inter and intra predictions) but also adds an additional prediction called the inter-layer prediction. This inter-layer prediction, processed by ELs only, allow to use coding information from the BL and lower ELs also called reference layers. This coding information include the reconstructed texture and the motion information and the coding mode. As a result, the inter-layer prediction allows bit-rate savings in the coding of ELs compared to simulcast.

However the inter-layer prediction also introduces an increase in the data storage needs. The coding information used for inter-layer prediction should be copied to the predicted layers. This copy is then signalled as a long term reference and is addressed with a specific reference index. It means the inter-layer prediction uses the HEVC prediction process. As a result, the inter-layer prediction process increases the HEVC complexity by adding a reference frame to predicted layers but does not add prediction tools.

Moreover, the copied coding information may have to fit the predicted layer format. As a result, additional time-consuming operations on the reference coding information may be processed before the inter-layer prediction. For instance, for spatial and High Dynamic Range scalability these operations are respectively up-sampling and inverse tone mapping. They also introduce operations ordering dependencies since they should be processed before the inter-layer prediction.

The inter-layer prediction implementation should then be considered as a trade-off between complexity and bit-rate savings. This trade-off can be applied at slice level.

B. Spatial scalability

In the case of spatial scalability, layers encode different resolutions. At most two additional operations are necessary for a scalable implementation. The first one enables the data to fit the quality level of the reference layer and is applied to uncompressed video. The second one enables the coding information to fit the predicted layer quality level. In the case of spatial scalability these operations are first a down-sampling and then an up-sampling filter. Only the up-sampling is defined in the SHVC standards and is applied at Coding Unit (CU) level. Luma and chroma components processed an 8-tap filter and a 4-tap filter, respectively. Compared to a single layer HEVC encoder, the spatial SHVC encoder complexity, is increased by both the up-sampling operations and the predicted layers encoding.

III. PROPOSED SHVC ENCODER

While the reference software SHM processes sequentially each layer, another architecture allowing parallel processing over layers can be considered. This architecture is based on the simulcast application where each layer can be processed at the same time. But, because of the inter-layer prediction, layers are not independent from each other and a synchronisation at layer level should be considered.

A. Initial HEVC encoder

The proposed SHVC architecture rely on a multi-support real-time software encoder developed by ATEME [17]. Video coding standards from MPEG-2 to HEVC are supported with a real-time coding rate. This performance is achieved thanks to parallelism at several level and Single Instruction Multiple Data (SIMD) on several operations. While used as an HEVC encoder, parallelism is performed over frame thanks to a pipeline architecture and over tiles [18] and over CTU lines in one slice per frame. In the rest of this paper, we refer to this real-time HEVC encoder as single layer encoder ATEME (SLE-A). The SLE-A encoding is based on a pipeline process at frame level. Pipeline operations are processed in independent processing threads. Each pipeline step ends with a synchronisation signal sending from the processing threads
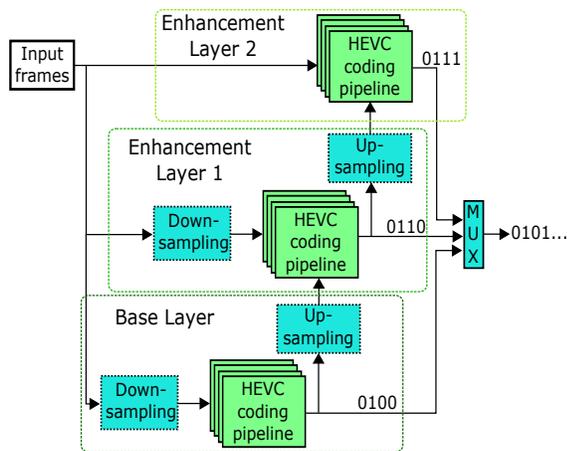
Fig. 1. SHVC-A architecture with spatial scalability

to the management thread. Once all synchronisation signals are received, the next step is launched. As a result, the cadence corresponds to the most time-consuming pipeline operation. The last pipeline step is the coding decision which produces the coding information that can be used for inter-prediction on other frames.

The SLE-A also integrates a bit-rate control module which allows to perform encodings at constant quantization parameter and Constant Bit-Rate (CBR). The proposed SHVC encoder inherits this bit-rate control module on each layer allowing them to allocate a dedicated encoding bit-rate in real-time. However this module is not considered in the rest of this paper since the Common Test Condition (CTC) [19] are defined under constant quantizer parameters.

B. SHVC proposed structure

The proposed SHVC encoder is composed of several SLE-A constituting the layers. The inter-layer synchronisation is performed at each pipeline steps. For a spatial scalability application, each SLE-A processes a different resolution and represents a spatial layer. In the rest of this paper, we refer to this SHVC encoder as SHVC-A.

In the proposed solution, the two additional down-sampling and up-sampling filtering operations constitute a pipeline step each on the reference layers as shown on Figure 1. These filtering operations are SIMD-optimised and parallelised over components plane and over data. These optimisations are necessary to reduce the encoding latency. Moreover, the inter-layer prediction can only be processed once the coding information from the reference layer is up-sampled. As a result, the reference layers pipeline length are increased by two steps while the predicted layers are delayed by three pipeline steps. Moreover, in common broadcast configuration, the B-slices are more numerous than other slice types. An up-sampling operation on each of these slice types would be highly time consuming. As a result, a trade-off is made on the slice types. The inter-layer prediction in SHVC-A uses I and P slices only as references.

This solution enables parallelism between layers and introduces a very low latency. Moreover, such architecture has a low impact on the SHVC complexity while increasing the number of layers.

The SHVC-A encoder can be used with two configurations for two purposes : real-time encodings and best coding efficiency encodings called SHVC-A Live and SHVC-A File, respectively. SHVC-A File configuration is designed to seek for the best coding efficiency with no collocated vectors. The SHVC-A Live configuration is designed for real-time encodings and proposes a trade-off between coding rate and coding efficiency on the ELs. This trade-off is achieved by adding limitations in the coding decision. For instance only one reference per list is allowed in the prediction process. The collocated vectors are not used. Prediction unit depth is limited to one. The SHVC-A Live configuration can be applied to all layers but is only applied to the EL. As a result, the BL is always set in the SHVC-A File configuration to seek the best coding efficiency.

IV. RESULTS

A. Experimental settings

The set-up used for experiment is the same as the one presented in our previous work [14], [15]. The platform employed is composed of four Intel Xeon E5-4627V3 CPUs. It provides 4x8 cores (32 threads) each running at 3.30 GHz.

The sequences used in the experiments presented are defined in the CTC. The experiments are conducted in a low-delay P-slices spatial $2\times$ configuration defined in the CTC [19]. In this considered configuration, the BL encodes I-slices and P-slices while the EL encodes P-slices and B-slices. This configuration is not commonly used in a broadcast set-up since it is more time consuming for encoders but here it allows to focus on the inter-layer prediction SHVC-A performs.

The performance of the proposed solution is assessed in terms of encoding frame rates for coding rate performance evaluation and bit-rate savings, measured with the Bjontegaard delta bit-rate (BD-BR) metric [20]. When only the SHVC EL is compared to the single layer equivalent, the BD-BR obtained corresponds to the bit-rate savings brought by SHVC over HEVC at the same video resolution and objective quality.

B. Coding rate

Table I exposes the coding rate in terms of frame per second achieved by the SHVC-A encoder in Live and File configurations. The presented speed-up are measured by dividing the SHVC-A coding rate by the considered encoder coding rate. Thanks to parallelism and optimisation and limitation in the coding decision employed in the EL, the SHVC-A is $2500\times$ and $100\times$ faster than the SHM in Live and File configurations, respectively. Moreover the SHVC-A Live encoder performs real-time encodings in most of the considered sequences, even with the low-delay configuration. The spatial scalability introduces less than 12% and 19% slow-down on encoding cadence compared to the initial encoder SLE-A in Live and File settings, respectively. Both up-sampling filter and

TABLE I
SHVC-A Average Coding Rate and Speed-up over SHM and
SLE-A in both Live and File configuration

Seq.	Coding rate SHVC-A (fps)		Speed-up SHVC-A over SHM		Speed-up SHVC-A over SLE-A	
	Live	File	Live	File	Live	File
A1	32.6	1.69	3397.8	177.3	0.90	0.81
A2	30.3	1.76	4172.8	242.5	0.92	0.90
B1	47.8	2.69	2680.0	150.5	0.91	0.88
B2	45.9	2.50	2552.9	139.2	0.88	0.83
B3	46.6	2.61	2585.9	145.6	0.89	0.86
B4	47.1	2.62	2849.3	159.0	0.90	0.87
B5	45.5	2.53	2520.0	141.7	0.88	0.85

TABLE II
Difference between Base Layer SHVC-A over SHM

Seq.	SHVC-A(BL) over SHM(BL)
A1	-10.9%
A2	-10.7%
B1	-9.1%
B2	-10.5%
B3	-12.1%
B4	-13.5%
B5	-18.1%
Avg.	**-12.1%**

the additional inter-layer reference considered in the coding decision are responsible for this slow-down. Since the Live configuration assumes limitations in the coding decision, the additional inter-layer reference has a lower impact on coding rate. As a result, the slow-down on encoding cadence in this configuration is less important.

These results shows that the proposed encoder is able to perform real-time encodings even with a low-delay configuration. They also show that the employed architecture, at worst, introduces only a 19% slow-down on encoding cadence compared to the basis encoder. This means the coding rate can be mostly preserved in a SHVC encoder implementation.
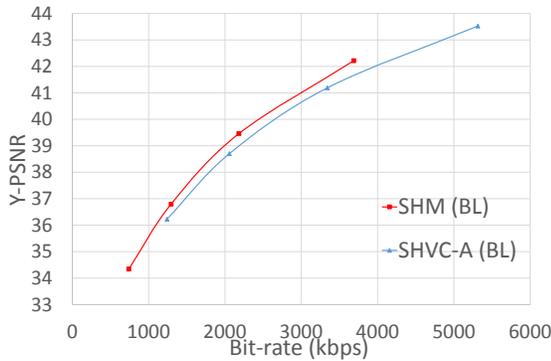
C. Bit-rate savings

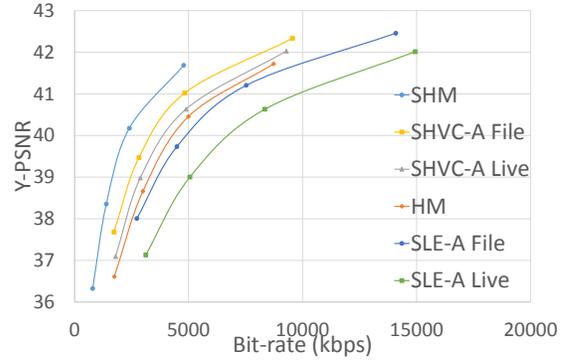Fig. 2. Y-PSNR vs Bit-rate on SHVC-A(BL) and SHM(BL) on B1 sequence

Fig. 3. Y-PSNR vs Bit-rate on SHM(EL) and SHVC-A(EL) and HM and SLE-A(EL) on B1 sequence

TABLE III
SHVC EL Bit-rate Savings over Simulcast

Seq.	SHM over HM	SHVC-A Live over SLE-A Live	SHVC-A File over SLE-A File
A1	-65.0%	-40.3%	-31.1%
A2	-28.3%	-51.8%	-36.5%
B1	-41.8%	-50.9%	-38.5%
B2	-51.4%	-32.6%	-25.6%
B3	-46.9%	-35.2%	-24.1%
B4	-38.1%	-36.3%	-21.2%
B5	-44.5%	-20.3%	-13.8%
Avg.	**-45.1%**	**-38.2%**	**-27.3%**

Figure 2 shows the rate-distorsion (RD) curves on the BL of the SHVC encoders on B1 sequence. SHM is better but the SLE-A is closed in terms of coding efficiency. Table II resumes these differences. On average SHM BL is more efficient by 12.1% over the SHVC-A BL.

Figure 3 shows the RD curves on the EL of the scalable and single layer considered encoders on B1 sequence. As expected, SHVC encoders performs bit-rate savings over their equivalent single layer encoder. SHVC-A performs less coding efficiency than SHM but still provides interesting bit-rate savings over SLE-A. Table III summarises these bit-rate savings. SHM performs better bit-rate savings than the proposed solution. This difference is explained by the difference between the

TABLE IV
Simulcast EL BD-BR Differences

Seq.	SLE-A Live over HM	SLE-A File over HM	SLE-A File over SLE-A Live
A1	61.3%	23.5%	-23.8%
A2	65.9%	12.0%	-32.0%
B1	56.3%	11.8%	-28.3%
B2	45.9%	17.9%	-19.3%
B3	71.0%	22.1%	-28.3%
B4	74.6%	17.2%	-31.9%
B5	88.8%	38.7%	-24.3%
Avg.	**66.2%**	**20.5%**	**-26.8%**

TABLE V
SHVC EL BD-BR DIFFERENCES

Seq.	SHVC-A Live over SHM	SHVC-A File over SHVC-A Live
A1	67.1%	-15.2%
A2	34.4%	-12.8%
B1	39.9%	-10.4%
B2	58.2%	-12.7%
B3	62.2%	-18.0%
B4	42.8%	-18.5%
B5	12.5%	-20.4%
Avg.	**45.3%**	**-15.4%**

BL coding efficiencies. The more coding efficiency the BL performs, the more bit-rate savings the inter-layer prediction enables. Furthermore, the SHVC-A Live performs more bit-rate savings than the SHVC-A File configuration. Since SHVC-A Live does not seek the best coding efficiency, it has more to benefit from the inter-layer prediction. Since SHVC-A EL does not use collocated vectors, it means only the reconstructed frame acts on the bit-rate savings on the SHVC-A.

Tables IV and V summarise the differences between the EL on the single-layer and scalable encoders, respectively. HM performs more coding efficiency than the proposed solution because of the limitations used on the File and Live configurations. Moreover the single layer Live configuration performs 26.8% less coding efficiency than the File configuration. This coding efficiency tendency is the same with the scalable encoders. However gaps are less important on the SHVC-A Live configuration. This is also due to the fact that the Live configuration benefit more from the inter-layer prediction. This means SHVC-A Live enables even more bit-rate savings over simulcast.

Finally, the proposed SHVC-A encoder performs interesting bit-rate savings even without using collocated vectors in the prediction. The proposed Live architecture benefits even more from the inter-layer prediction than the File set-up does.

V. CONCLUSION

In this paper, we have presented a $2\times$ spatially scalable software SHVC encoder achieving encoding bit-rate savings over simulcast at real-time coding rate. The proposed architecture allows parallelism over layer and introduces a cadence slow-down inferior to 19% of the simulcast equivalent total coding rate. Experiments show the proposed solution performs real-time encodings on 1080p60 and 1600p30 resolution in a low-delay set-up thanks to multiple parallelism employed and the limitations exposed in the Live configuration. The bit-rate savings over simulcast are less important than state of the art but remains interesting in a real-time set-up. The experiments also show that the more coding efficiency the BL performs, the more bit-rate savings the inter-layer prediction performs. Moreover, UHD contents real-time encodings were demonstrated with the proposed solution in a random-access configuration at an ATSC 3.0 meeting [21] for which SHVC is a video coding solution candidate [13].

As future work, the authors will investigate on down-sampling and up-sampling filtering to achieve spatial scalability encodings with more ratio than only $2\times$. Moreover authors already study the inter-layer prediction indirect impact when more than one EL are used and also bit-rate savings and coding rate impact when B-slices are used as inter-layer reference.

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] 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, pp. 560–576, Jul. 2003.

[4] 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.

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

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

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

[11] X. HoangVan, J. Ascenso, and F. Pereira, "Improving SHVC Performance with a Joint Layer Coding Mode," *IEEE International Conference on Acoustics, Speech and Signal Processing*, march 2016.

[12] T. Biatek, W. Hamidouche, J. F. Travers, and O. Deforges, "Optimal bitrate allocation in the scalable hevc extension for the deployment of uhd services," *IEEE Transactions on Broadcasting*, vol. 62, no. 4, pp. 826–841, Dec 2016.

[13] "A/341: ATSC Candidate Stadard: Video - HEVC," Approved 24 November 2015. Updated 3 January 2017. The CS period ends 30 January 2017.

[14] R. Parois, W. Hamidouche, E. Mora, M. Raulet, and O. Deforges, "Efficient Parallel Architecture of an Intra-only Scalable Multi-Layer HEVC encoder," *Conference on Design & Architectures for Signal & Image Processing (DASIP)*, october 2016.

[15] ——, "Real-time UHD Scalable multi-layer HEVC encoder architecture," *24th European Signal Processing Conference*, september 2016.

[16] W. Hamidouche, M. Raulet, and O. Deforges, "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.

[17] ATEME, "http://www.ateme.com."

[18] 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.

[19] 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.

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

[21] 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.