Reducing Computational Complexity in HEVC Decoder for Mobile Energy Saving

Naty Sidaty, Julien Heulot, Wassim Hamidouche, Erwan Nogues, Maxime Pelcat and Daniel Menard
INSA Rennes, IETR Lab, CNRS 6164
Email: firstname.lastname@insa-rennes.fr

***Abstract*—With the growing development of video applications and services for mobile devices, saving energy consumption when managing video is becoming a more and more important issue. The challenge is then to deliver video with high quality while reducing the energy consumption. In this paper, we investigate the relationship between subjective video quality and energy consumption in an HEVC decoder. By reducing the computational complexity of the decoder, drastic energy savings can be achieved without affecting the visual quality. In this paper, two computation methods and several filter configurations are tested. Results show that at least 10% of energy savings are obtained with the same subjective perceived quality. In addition, objective measurements have shown that only a slight quality degradation has been noticed.**

I. Introduction

Today, with the increase in video communication services, social networking and user requirements to watch videos anytime, anywhere and on various mobile devices, the duality energy consumption and video quality has become an active research topic. Over the past two decades, several video coding tools have been developed in order to ensure a high Quality of Experience (QoE) for such services. The latest video coding standard High Efficiency Video Coding (HEVC) [1] enables up to 60% [2], [3] bit rate reduction, for the same perceived quality, with respect to the Advanced Video Coding (AVC) standard [4]. Moreover, the Scalable HEVC (SHVC) [17] was proposed one year and half later to provide spatial, color gamut, quality and bit-depth scalability with a coding gain up to 30% compared to simulcast HEVC coding. On the other hand, a huge effort has been devoted, whether at circuit or system levels, in order to reduce computational energy consumption. From one side, technologies-based circuit, enable shrinking down transistors and, consequently, providing more energy efficient hardware. From the other side, technologies-based systems are developed to adapt the instantaneous processing capacity to the requirements of the running applications [7]. Dynamic Voltage Frequency Scaling (DVFS) and Dynamic Power Management (DPM) methods [8] allow reducing energy consumption without modifying an application's behavior. Thus, no significant influence on data processed is noticed when deactivating them.

The proliferation of connected devices (smartphones, tablets, etc.) further stresses the need to carefully design algorithms, reducing encoding/decoding video complexity process to save energy consumption while ensuring a high quality, required by end users. This later can be measured either by subjective or objective quality metrics [9]. Subjective quality assessments are carried out by involving human viewers assessing the perceived quality. They are the most reliable methods for assessing the quality of a multimedia service because the human eye is often the last point in the multimedia transmission chain. The objective metrics are, on the contrary, signal-based measures without taking into account the human visual system properties.

The low power HEVC decoding systems have been the object of different studies [10], [11]. The energy consumption saving can mainly obtained by reducing the decoder computational complexity. These decoder optimisation's process may cause video quality degradations despite the computational energy gain. There is, therefore, a real compromise between video quality and energy consumption.

II. Related Works

"Fig. 1" represents the block diagram of HEVC standard decoder. Input compressed data are processed firstly in the entropy decoder block by extracting the different syntax elements (SE). The residuals are then inversely quantized and transformed in the Inverse Quantization & Inverse Transform block. The blocks prediction is applied in intra/inter-frame based on input parameters in the data bitstream [1]. This prediction is based on the previously decoded image (in inter-frame case). Motion vectors, having a fractional pixel resolution, are also transmitted in the bitstream data, as depicted in "Fig.1". Finally, in-loop filters are applied on the decoded picture to avoid potential artifacts and enhance the image quality.

These decoding process have been profiled on different platforms at various encoding configurations. Two platforms have been used: General Purpose Processor (GPP) and Digital Signal Processor (DSP) [12], [13]. In addition, under common test conditions [14], two coding configurations are considered: Random Access (RA) and Low Delay (LD). RA configuration, used generally for broadcasting applications, is based on pyramidal structure using Intra (I), Predicted (P) and Bi-predicted (B) frames. A picture of a temporal sub-layer i is predicted using picture of lower temporal sub-layers j with $(j < i)$. LD configuration is, on the contrary, based on horizontal structure (IPPP), where a given frame of a sub-layer i is predicted based on the I-frame of the same sub-layer.

ISBN 978-0-9928626-7-1 © EURASIP 2017

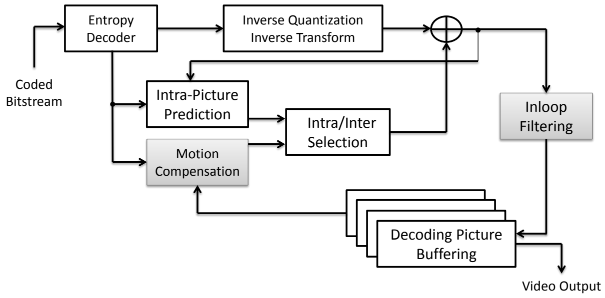


Fig. 1: Diagram of HEVC standard decoder.

The HEVC standard decoder optimisations have been the object of numerous studies [15], [16]. A weak point of the decoder is the relative complexity of data processing blocks. “Table I” presents an example of complexity analysis per processing block of the HEVC decoder, using *kimono1* sequence (1920×1080) in RA configuration, with different quantization parameters (QP).

Motion Compensation (MC) and In-loop Filtering blocks,

TABLE I: Complexity analysis per processing blocs on HEVC decoder.

QP	22	24	26	28	30	32	**Avg**
MC filters	62.5	69.9	72.3	75.3	78.1	80.2	73.7
Inv. Transform	13.2	11.6	9.5	8.7	8.0	7.4	9.3
In-loop filtering	13.9	10.0	10.5	9.2	7.8	7.1	9.6
Entropy decod.	6.4	5.0	4.2	3.6	3.3	2.2	4.2
Others	4.0	3.5	3.5	3.2	2.8	3.1	3.2

depicted in gray color in “Fig. 1”, require together more than 83% of computation resources. This is due to the fact that MC and in-loop filters are used to treat the uncompressed data stream, compared to the other blocks where only compressed data are treated. Therefore, reducing the complexity of these two block treatments could be benefit for decoding energy saving. Our proposal is based on filter complexity reducing and two main techniques are considered here: Computation Approximation and Computation Skipping.

A. Computation Approximation

This technique consists to replace a complex processing block by a set of simplified blocks. This is done by using different parameters instead of a complex algorithm or replacing a complex operator by another of a lower complexity. “Fig. 2” depicts the MC filters approximation by replacing the MC filter by a set of low complex filters, where legacy filters are replaced by approximate filters. Using an *approximation level control*, three filter categories are created, based on the number of the filter taps: low, intermediate and middle, as given in “Table II”. The Legacy corresponds to the original filter taps in the HEVC standard decoder.

B. Computation Skipping

For the same purposes, this technique is used in order to disable some additional processing parts for reducing the

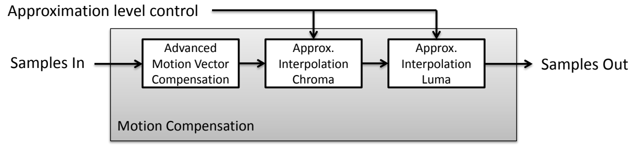


Fig. 2: Interpolation filter of the MC block approximation.

TABLE II: Filter size per configuration (number of taps).

Configuration	Chrominance filter size	Luminance filter size
Low	1	1
Middle	2	3
Intermediate	3	7
Legacy	4	8

computational complexity. The post and pre-processing blocks, used for signal enhancements, can be skipped (periodically or regularly) without affecting the video quality. The SAO filter, in “Fig. 3”, is used to reduce sample distortion with a specific reconstructed sample classifications, using different offset values [16]. A *Skip Control* parameter is used to classify the video distortion at the decoder side. Thus, a decision is taken to activate or not the in-loop filters. “Table III” shows the percentage of block skipping for different filter categories, based on activation in-loop filter frequencies. The Legacy corresponds to a *Skip Control* $= 0$. In the rest of the paper, these three filter configurations in both computational methods (Approximate and Skipping) and contexts (RA and LD) are adopted.

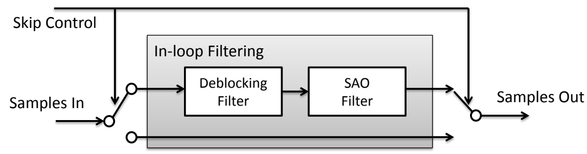


Fig. 3: In-loop filter skip.

TABLE III: Percentage of block skipping per configuration.

Configuration	Block skipping percentage
Low	89%
Middle	63%
Intermediate	25%
Legacy	0%

In this paper, we propose to investigate the influence of these processing-oriented (computation) methods on perceived video quality and energy consumption by caring out a set of subjective studies on a mobile platform. These methods allow to decompose an application into several blocks with a low computation complexity. By doing this, a drastic energy saving and an unaffected video quality can be achieved.

The rest of the paper is organised as follow: Section II introduces the context and related works. In Section III we detail the conducted quality assessment experiments, where different filters and computation methods are tested. Results

ISBN 978-0-9928626-7-1

and associated discussions are presented in Section IV. Finally, conclusions and future works are given in Section V.

III. Quality Assessment Experiments

In order to investigate the influence of processing-oriented methods on perceived quality, we have conducted a set of psycho-visual experiments. Bellow, we describe the experimental environment, the used video database and the evaluation procedure.

A. Experimental Environment

Three subjective quality assessment experiments have been carried out. The first experiment is based on computation skipping (Skip), the second on computation approximation (Approx) and the last experiment is based on comparison between Low delay P (LD.P) and Random Access (RA), as shown in "Table IV". All these tests have taken place, separately, in our psycho-visual room complying with the ITU-R BT.500-13 Recommendation [18]. A display 9-inch universal HDMI tablet 1280 x 800, with 10-points capacitive multi-touch screen (ODROID-VU), was used to visualise the video sequences.

TABLE IV: Experiment's Configurations

Experiment	Computing Techniques	Configuration	QP
1	Skip	RA	27 & 32
2	Approx	RA	27 & 32
3	Skip & Approx	LD.P & RA	27

B. Test Sequences and Configurations

A set of video sequences, from different categories, has been selected from MPEG database[1]. The choice of these videos is mainly based on the color content, movement and texture. Five original sequences, of 10 seconds duration each, have been used in the experiment, as given in "Table V". An example frame of the used sequences is given in "Fig.4". These sequences are processed with the HEVC encoder/decoder reference software. Different configurations are then tested (computation skipping, computation approximation, with LD and RA). Two quality levels, based on Quantization Parameter (QP), are considered: QP=27 and QP=32. Finally, we define three filter categories, adapting their computational complexity: low, middle, intermediate, and legacy (reference), respectively [7].

C. Participants

Sixteen healthy naive subjects have been invited to participate to these experiments. There were 11 male and 5 female with ages ranging between 21 and 43. Visual acuity has been checked by FrACT (Freiburg Visual Acuity Test) and it was around 1 with or without correction. All subjects have been screened using the Ishihara compatible color vision test for detecting color blindness. In order to ensure reliable scores, all participants have been gratified.

[1]This dataset is used by JCTVC MPEG Team in HEVC test conditions for compression

Fig. 4: Snapshots of the used video sequences.

TABLE V: Test video sequences

Sequence	Duration (sec)	Resolution	Fps
Cactus	10	1280 x 720	50
BQTerrace	10	1280 x 720	60
BasketballDrive	10	1280 x 720	50
ParkScene	10	1280 x 720	24
Kimono1	10	1280 x 720	24

D. Evaluation Procedure

In our subjective quality assessments, the Double Stimulus Continuous Quality Scale (DSCQS) method was used [18]. Each video sequence was presented twice to participants accompanied by its reference version (original). Participants were asked to judge the quality of presented stimulus, according to a rating scale, given in "Table VI". They have 10 seconds during to give and confirm their quality scores. Collected data, from three experiments, have been treated separately but applying the same treatments.

TABLE VI: Example of rating scale used in the experiments.

Degradations	Quality	Score
Imperceptible	Excellent	5
Perceptible But Not Annoying	Good	4
Slightly Annoying	Fair	3
Annoying	Bad	2
Very Annoying	Poor	1

IV. Results and Discussion

Two quality measures have been done in order to investigate the influence of processing-oriented methods on perceived quality: Subjective evaluations and objective measurements. In addition, energy consumption-based quality is then analysed.

A. Subjective Assessments

Subjective results and analysis in this section are based on Mean Opinion Score (MOS), for each video used in each experience. This is given by (1).

$$MOS_{jk} = \frac{1}{N}\sum_{i=1}^{N} s_{ijk} \quad (1)$$

where s_{ijk} is the score of participant i for modified sequence j of the video k and N is the number of participants.
Generally, these values are associated with the corresponding confidence intervals, usually at 95%. The smaller the values,

the more reliable are the results. “Fig. 5 (a)” shows the obtained subjective quality scores for two computation methods (Approx & Skip) on the average sequences. Results show very limited quality degradations compared to the reference filter (legacy), regardless the used QP. Moreover, participants have better judged, on the average, the quality perceived with “inetrmediate” filter than with the “legacy” reference filter.

B. Objective Measurements

Peak Signal-to-Noise Ratio (PSNR) is commonly used in image and video quality assessment as objective metric. This is done by weighting three components Y, U and V as bellow:

$$wPSNR = (6 * PSNR_Y + PSNR_U + PSNR_V)/8 \quad (2)$$

This signal-based metric does not fully represent the quality of degraded video as they can be perceived by human eyes. Another widely used metric is the Similarities quality (SSIM). The evaluation processes are made on image structures, instead of pixels, as done in PSNR. “Fig. 5 (b) and (c)” depict the average of all used sequences, using $PSNR$ and $SSIM$ metrics, respectively. Using two computation methods (Approximate & Skipping), results show only a slight quality distortions have been noticed, apart for “Low” filter configuration. In other words, the proposed methods performance lead to a low quality distortions compared to the legacy decoder, especially for computation skipping.

C. Energy Consumption

As noticed above, the proposed-oriented methods allow a computation complexity reduction without affecting the perceived quality. For energy measurements, an octa-core Exynos 5410 SoC has been used. This System on Chip (SoC) is based on the big.LITTLE configuration including a cluster of 4 ARM Cortex-A15 cores and a cluster of 4 ARM Cortex-A7 cores. Power sensors are embedded in this SoC allowing power measurements during the processing. For these experiments, power measurements are done after each frame processing and averaged over the complete video decoding process. The computation is done only on the 4 ARM Cortex-A15 cores. Thus, the thread scheduling has no impact on consumption measurements. Finally, the standard "OnDemand" DVFS governor from Linux have been used to reduce static power consumption during the experiments.

In “Fig. 6 (a)” the average energy consumption of all used sequences is presented. As we can noticed, a drastic energy saving can be obtained by the computation configurations. In computation approximate, for example, at least 10% energy saving $(0.1W)$ can be easily achieved by using the use of “Middle” filter instead of the reference HEVC decoder filter (legacy). In fact, the two filters have the same perceived quality, as given in “Fig. 5 (a)”, but with different energy consumption. In Skipping computation, this gain is slightly greater than in Approximate computation, apart for “Low” filter configuration.

A trade-offs between energy consumption and quality are presented in “Fig. 6 (b) and (c)”, for Approximate computation. Using subjective results and objective $wPSNR$ measures, respectively, results show that using modified filters (intermediate and middle) allow, subjectively, a considerable energy saving with high perceived quality than that obtained in legacy reference filter. In addition, these behaviours are confirmed using quality degradation measures, $wPSNR$, where only a slight quality differences are noticed between “intermediate” and reference HEVC decoder filter. It is interesting to note here that all these results are obtained in Random Access (RA) context.

D. Random Access vs Low Delay P

In this section, we have further investigated our proposed methods by introducing a Low Delay P (LD) context. “Fig. 7 represents a performance comparison between RA and LD, based on subjective results (a), objective measures (b) and energy consumption (c). Results demonstrate that, in LD context, video quality is significantly affected compared to the RA. On the contrary, a considerable energy saving can be achieved. Finally, in this context, as in the results above, computation “Skipping” allows a better quality and energy saving that in computation “Approximate”. The Group Of Picture (GOP) in LD coding configuration propagates the “Approximate” and “Skipping” errors in the video more than in the RA coding configuration.

V. Conclusion

In this paper, a set of psycho-visual experiments have been performed in order to study the compromise between mobile video quality and decoder energy consumption. Comparing to the complexity of a standard HEVC decoder, we have investigated the use of two approximate computing methods: “skipping” in-loop filters and “approximating” the motion filters. Two context studies (RA & LD) and different categories of tests have been conducted. Experimental results show that with both approximate and skipping methods, drastic energy savings can be achieved. In addition, no significant quality degradation is noticed when filters are modified. Opposite properties are observed in the RA and LD contexts. In fact, for the same configuration, RA offers a better perceived quality while the LD coding configuration makes it possible to save more energy. In future work, we look to incorporate the new HEVC coding tools, as JEM, and the emerging video applications, such as Virtual Reality (VR, 360’), with higher image resolutions (4K, 8K).

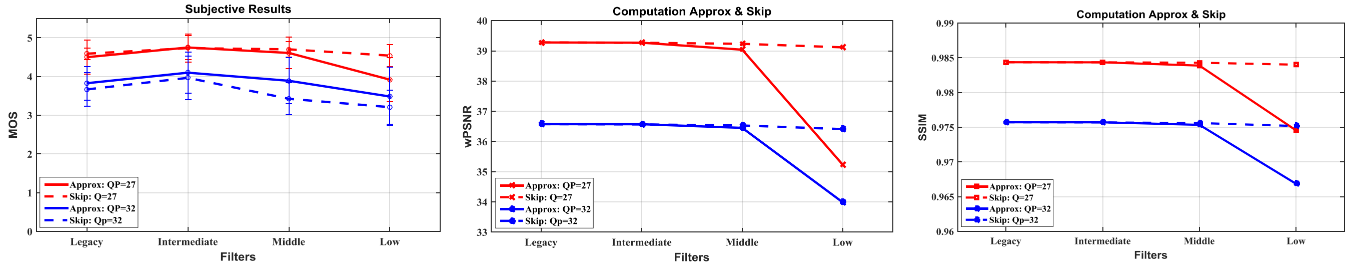
References

[1] J-S. Gary, 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.

[2] R. Ohm, G. Sullivan, J. 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, 2012.

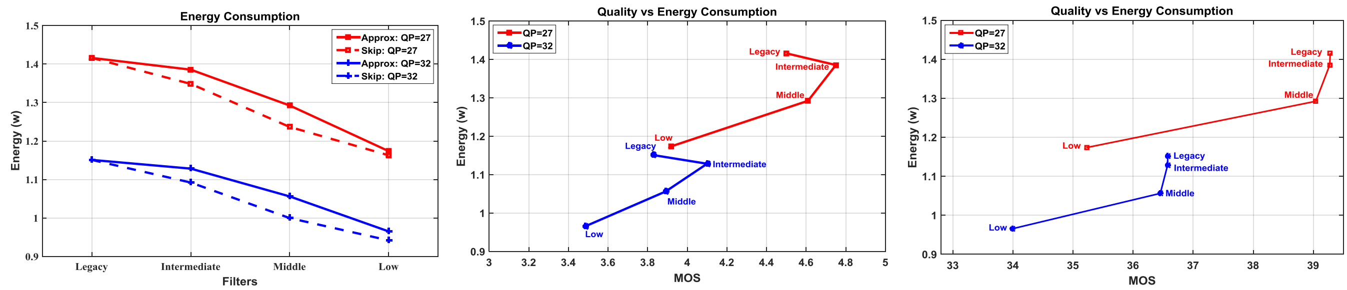
[3] T-K. Tan, R. Weerakkody, M. Mrak, N. Ramzan, V. Baroncini, J-R. Ohm, and J-S. Gary, *Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance,* IEEE Transactions on Circuits and Systems for Video Technology , vol. 26, no. 1, pp. 76-90, Jan. 2016.

ISBN 978-0-9928626-7-1 © EURASIP 2017



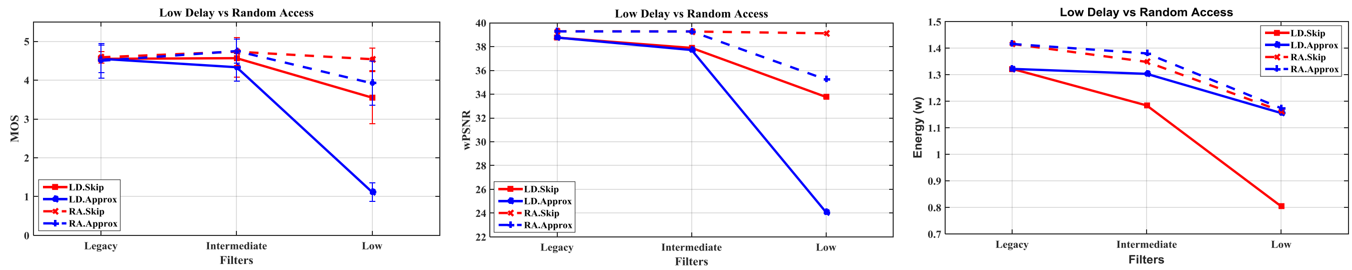
(a) *Subjective quality.* (b) *wPSNR measures.* (c) *SSIM measures.*

Fig. 5: Subjective and objective results for two experiment configurations: in Approx & in Skip.



(a) *Energy consumption per configuration.* (b) *Subjective quality **vs** energy consumption.* (c) *Objective quality **vs** energy consumption.*

Fig. 6: Trade-offs between quality and energy consumption.



(a) *Subjective quality* (b) *Objective measures* (c) *Energy consumption*

Fig. 7: Random Access (RA) *vs* Low Delay P (LD).

[4] T. Wiegand, G.J. Sullivan, G. Bjøntegaard, and A.Luthra, *Overview of the H.264/AVC video coding standard,* IEEE Tran. on Circuits and Systems for Video Technology ,vol. 13, no. 7, pp. 560-576, July 2003.

[5] A. Sachid, R. Francis, M. Baghini, D. Sharma, K-H. Bach, R. Mahnkopf, and V. Ramgopal Rao, *Sub-20 nm gate length finfet design: Can high-spacers make a difference ?* in IEEE International Electron Devices Meeting, IEDM'08, pp. 1-4, Dec. 2008.

[6] B-Y. Nguyen, C. Mazure, D. Delprat, C. Aulnette, N. Daval, F. Andrieu, and O. Faynot, *Overview of fdsoi technology from substrate to device,* in International Semiconductor Device Research Symposium, ISDRS'09, pp. 1-2, Dec. 2009.

[7] E. Nogues, D. Menard, and M. Pelcat, *Algorithmic-level Approximate Computing Applied to Energy Efficient HEVC Decoding,* IEEE Transactions on Emerging Topics in Computing, vol. PP, no. 99, pp. 1-1, 2016.

[8] L. Benini and G. DeMicheli, *Dynamic power management: design techniques and CAD tools,* Springer Science & Business Media, 2012.

[9] M. Pedram, Abbas E-M, S. Shirani, *ubjective and Objective Quality Assessment of Image: A Survey,* Computer Vision and Pattern Recognition, Elseiver 2014.

[10] C. C. Chi, M. Alvarez-Mesa and B.Juurlink, *Low-power high-efficiency video decoding using general-purpose processors,* ACM Trans. Archit. Code Optim., vol. 11, no.4, pp. 56:1-56:25, Jan. 2015.

[11] E. Raffin, E. Nogues, W. Hamidouche, S. Tomperi, M. Pelcat, and D. Menard, *Low power hevc software decoder for mobile devices,* Journal of Real-Time Image Processing, pp. 1-13, 2015.

[12] F. Bossen, B. Bross, K. Suhring, and D. Flynn, *Hevc complexity and implementation analysis,* Circuits and Systems for Video Technology, IEEE Transactions on, vol. 22, no. 12, pp. 1685-1696, 2012.

[13] W. Hamidouche, M. Raulet, and O. Deforges, *Parallel shvc decoder: Implementation and analysis,* IEEE conference on ICME, 2014.

[14] F. Bossen, *Common Conditions and Software Reference Configurations,* Document JCTVC-H1100, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, San Jose, CA, Feb. 2012.

[15] M. Chavarrias, F. Pescador, M. Garrido, M. Raulet et al., *A dsp-based hevc decoder implementation using an actor language dataflow model,* Consumer Electronics, IEEE Tran. on, vol. 59, no. 4, pp. 839-847, 2013.

[16] C-M. Fu, E. Alshina, A. Alshin, Y-W. Huang, C-Y. Chen, C-Y. Tsai, C-W. Hsu, S-M. Lei, J-H. Park, and W-J. Han, *Sample adaptive offset in the hevc standard,* Circuits and Systems for Video Technology, IEEE Transactions on, vol. 22, no. 12, Dec. 2012.

[17] Jill M. Boyce, Yan Ye, Jianle Chen and Adarsh K. Ramasubramonian, *Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard*, IEEE TCSVT, vol. 26, no. 1, Jan. 2016.

[18] ITU-R BT.500-13 Recommendation, *Methodology for the subjective assessment of the quality of television picture,* Geneva, 2012.