An overview of the emerging HEVC standard

Philippe Bordes+, Gordon Clare$, Félix Henry$, Mickaël Raulet*, Jérôme Viéron§
+Technicolor, $Orange Labs, *IETR/INSA Rennes, §Ateme
France
philippe.bordes@technicolor.com; gordon.clare@orange.fr; felix.henry@orange.com; mraulet@insa-rennes.fr; j.vieron@ateme.com

***Abstract*—This paper presents an overview of the upcoming High Efficiency Video Coding standard (HEVC). Among all HEVC requirements, HEVC has to divide by a factor of 2 the bitrate of the current state-of-the-art standard H.264/AVC. At the 8th standardization meeting, the joint collaborative team (JCT) has reached the goal and even more has reduced the bitrate up to 70% compared to AVC/H.264 at constant quality (psychovisual tests). This paper will present a state-of-the-art of current coding and parallelism tools available in the HEVC standard.**

Keywords: High Efficiency Video Coding, Video Coding

I. Introduction

A new video compression standard, known as High Efficiency Video Coding (HEVC) [1], is currently under its final standardization stage and has been developed by a Joint Collaborative Team of ISO/IEC MPEG and ITU-T VCEG (JCT-VC). The Final Draft International Standard is expected to be delivered in January 2013. The overall requirements of HEVC is to improve the compression efficiency by a factor of at least two compared to the H.264/AVC compression standard, H.264/AVC having the best coding efficiency performance among the current generation of video standards. Since the first joint meeting in April 2010 where 27 proposals were evaluated in comparison with the reference software of H.264/AVC, there is a growing interest of semiconductor and mobile phone companies around this new video compression standard. From the top 3 proposals, achieved PSNR results were good enough (around 35%) to launch the HEVC standard activity. Both objective (PSNR-based) and subjective quality assessments have been performed at the 7th meeting within the JCT-VC and test results confirm that the initial goal can be reached:

- In its best encoding profile, using the full set of tools, HEVC can provide a bit rate savings around 36% for equal PSNR for the 1080p test sequences [2],
- The bitrate savings when considering equal subjective quality is even greater, more than 50% for all the test sequences and up to 70% for two of them [3], [4].

Moreover, despite the increasing complexity of Intra and Inter prediction modes, the encoding and decoding computation costs seem to be very well mastered, since:

- The encoding time is increased only by 10% and the decoding time by 60% in its best encoding profile and coding structure, compared to H.264/AVC reference SW model [4].

Even if one can argue that the HEVC reference Software model is written in a most efficient way than the H.264/AVC reference Software model, this limited extra computation time is good news to allow real-time coding and decoding of bigger video formats (bigger picture resolution, higher dynamic ranges and/or frame rates) in the next few years.

The rest of the paper will be divided as follows: Section II will briefly introduce the coding tools in the HEVC standard. Section III will describe one of the biggest evolution of HEVC is its ability to provide a bitstream that can be parallelized.

This work is done as part of 4EVER, a French national project with support from Europe (FEDER), French Ministry of Industry, French Regions of Brittany, Ile-de-France and Provence-Alpes-Cote-d'Azur, Competitivity clusters "Images&Reseaux" (Brittany), "Cap Digital" (Ile-de-France) and "Solutions Communicantes Sécurisées" (Provence-Alpes-Cote-d'Azur).

978-1-4244-5997-1/10/$26.00 ©2010 IEEE

II. Coding tools for HEVC

HEVC is a hybrid video codec. Some existing tools from H.264/AVC have only been revisited in this standard. HEVC has been designed to target ultra high resolution with higher framerates compared to H.264/AVC. Taking this into consideration, HEVC has introduced a new partitioning image scheme concept based on a quadtree structure with larger block size – a 64x64 Coding Unit (CU). A Coding Unit can be recursively divided into 4 CUs (Quadtree). Optionally, all the samples based processing and the reference pictures storage may be made using 10 bits precision (Internal Bit Depth Increase, IBDI).

A. Intra coding tools

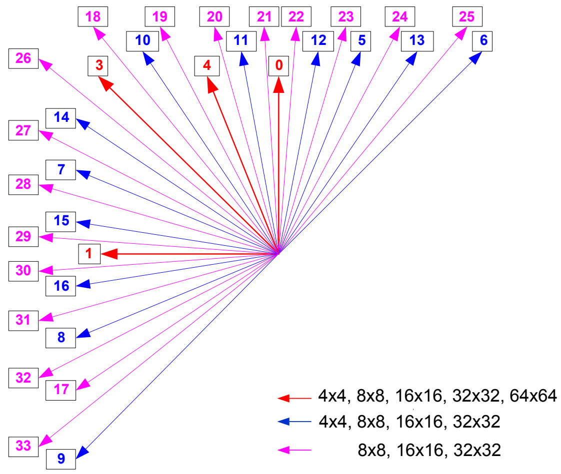

Figure 1: Intra prediction directions.

The number of Intra prediction directions depends on the CU size as depicted in Figure 1: 33 angular modes for 8x8, 16x16 and 32x32, 16 for 4x4 and 4 for 64x64 (LCU). As for H.264/AVC, the DC mode corresponds to the mean value of the samples from both top row and left column and I_PCM allows for coding sample values directly. Additionally, two new intra modes have been introduced: the planar mode (depicted in Figure 2) and the LM Chroma mode, where the chroma samples are predicted from the reconstructed luma samples

The top and left edges of the Intra DC and angular predictions are filtered. The number of filter taps depends both on the intra direction and the CU size (Mode Dependent Intra Smoothing).

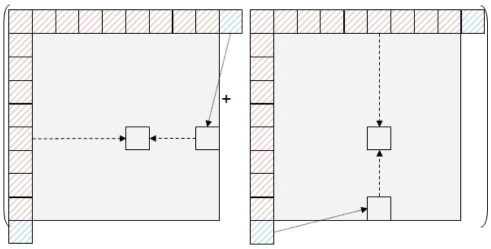

Figure 2: Planar mode: Interpolation indicated by dashed arrow, replication indicated by dotted arrow.

The decoding of the Intra mode is made in 2 steps. First the 3 Most Probable Modes (MPM) are derived from the neighboring coded CUs. Next, a decoding mode tree allows to derive the intra mode from one binary flag, and 1 or 2 more syntax elements.

B. Inter coding tools

A non-split CU is a Prediction Unit (PU) which can be sub-partitioned into 4 square or 2 rectangular partitions (Figure 3). In the same way as H.264/AVC, each PU partition is built with uni- directional or bi-prediction motion compensation, using ¼ (luma) or 1/8 (chroma) pel precision motion vectors (*mv*). But the *mv* values are predicted using motion vector competition: an index corresponding to a list of (spatial and temporal co-located) mv predictors is coded (AMVP).
The modes Merge and Skip (no residuals) enable deriving *mv* and reference index from 1 or 2 neighboring CUs.

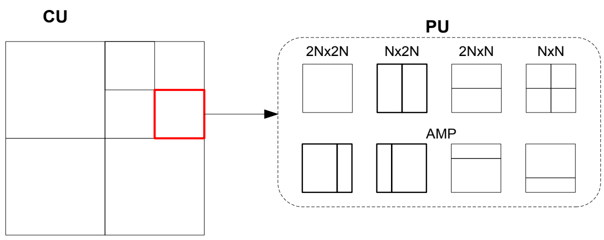

Figure 3: CU and PU partitioning.

C. Residuals coding

For transform coding of the prediction residual, a CU can be split into smaller square transform units (TU). The splitting is signaled using a second quadtree, the residual quadtree (RQT) [5]. RQT allows to adapt the transform to the frequency characteristics of the residual signal. The Transforms are DCT 4x4, 8x8, 16x16 or 32x32, except for Intra 4x4 that is DST (Mode Dependant Decoding Transform). In case of Asymmetric Motion Partitioning (AMP) (Figure 3), Non-Square Transform (NSQT) is used.

The Coding Unit Quantization Group (CUQG) is specified as a superset of CU for conveying Quantization Parameter (QP) values. Each CUQG is composed of single or multiple CUs, all of which have same QP values. Average QP of left and top CUQG (if available) are used as predictor for current CU quantization parameter.

The significance map (SM) of a TU indicates the positions of non-zero coefficients in the TU. For the largest TU size, a two level structure allows for splitting the SM into 4x4 SMs for coding.
In Intra, the scanning order of the coefficients (zig-zag, horizontal or vertical) is selected depending on the mode (Mode Dependent Coefficient Scanning, MDCS).

The first coefficients (levels) are coded using CABAC, next Golomb Rice and the last levels with Exp. Golomb. For each 4x4 coefficient set that satisfies certain condition the sign bit of the first nonzero coefficient is not coded but later inferred from the parity of the sum of all nonzero coefficients in that set.

D. Filtering SAO, ALF, Deblocking

In order to restore the degraded frame caused by compression, three kinds of filtering are applied successively. First a deblocking filter reduces the blocking artifacts in the same way as H.264/AVC. The smoothing strength depends on the QP value and on the reconstructed sample values difference at the CU boundaries.

The Sample Adaptive Offset (SAO) classifies reconstructed pixels into categories and reduces the distortion by adding an offset to pixels of each category in current region.

Band offset (BO) classifies all pixels of a region into multiple bands where each band contains pixels in the same intensity interval. The intensity range is equally divided into 32 intervals from zero to the maximum intensity value (e.g. 255 for 8-bit pixels), and each interval has an offset. Next, the 32 bands are divided into two groups. One group consists of the central 16 bands, while the other group consists of the rest 16 bands. Only offsets in one group are transmitted.

Edge offset (EO) uses four 1-D 3-pixel patterns for pixel classification with consideration of edge directional information, as shown in **Figure 4**. Each region of a picture can select one pattern to classify pixels into multiple categories by comparing each pixel with its two neighboring pixels.

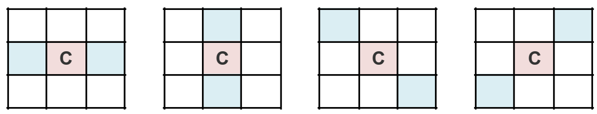

Figure 4: EO pixel classification patterns.

At last, an Adaptive Loop Filtering (ALF) allows to minimize the MSE between the reconstructed and source frames. It classifies the pixels into multiple categories based on the local directional characteristics. A Wiener filter is estimated and transmitted for each category by minimizing the average mean square error between the original and reconstructed pixels in that category. ALF can be applied to the entire frame or to local areas.

E. High level syntax

As for H.264/AVC, the video bitstream is composed of Network Abstraction Units (NALUs). HEVC introduces 4 NALU types for coded slices: Instantaneous Decoder Refresh (IDR) picture; a Clean Random Access (CRA) picture; Temporal layer access (TLA) picture; and a non-IDR, non-CRA and non-TLA picture. CRA is used to signal open Group of picture (GoP). Compared to H.264/AVC wherein a CRA picture may be signaled using a recovery point Supplemental Enhancement Information (SEI) message, in HEVC a distinct NAL unit type is used to indicate a CRA picture. Temporal layer access (TLA) picture is introduced in HEVC to indicate temporal layer switching point for temporal scalability.

In addition to coded slices, Sequence Parameter Sets (SPS) are used to carry data valid to the whole video sequence, whereas Picture Parameter Sets (PPS) carry information valid on a picture-by-picture base. A new NALU has been added in HEVC compared to H.264, the Adaptation Parameter Sets (APS) carrying picture-adaptive information that is also valid on a picture-by-picture base but is expected to change more frequently than the information in PPS.

Unlike its predecessors, HEVC focuses on progressive content coding only. However, legacy interlaced content representing still a very large part of the current video traffic, it has been decided to provide a free support (without core-design change) of interlace through the adoption of the Field Indication SEI message which indicates source type and some relative information for field applications.

F. HEVC performance

[2] presents the results of objective (PSNR-based) compression comparison tests between the current state of the HEVC draft standard and the AVC High Profile as an anchor reference. The performance is measured with all HEVC tools activated except 10-bit encoding. The following results are obtained:

- For the all-intra configuration, HM-6.0 (HEVC reference software) can save about 24% in bit rate;
- For the random-access configuration, HM-6.0 can save about 36% in bit rate; and
- For the Main Profile low-delay configuration, HM-6.0 can save about 37% in bit rate.

However, coding performance increases for larger resolutions. For instance, in random access configuration, and for HD and beyond the gain in compression is larger than 40%. Furthermore, first subjective tests [3],[9] seems to indicate the HM encoder is saving at least 50% bit rate compared to JM for equivalent perceived quality.

III. High level parallelism tools

A. Slices

In HEVC, it is possible to divide a frame into slices, as in H.264/AVC. Slices are groups of LCUs in scan order. Slice can be used both for network packetization and for parallel processing. However, a severe penalty on rate distortion performance is incurred when using slices, due to the breaking of all dependencies at their boundaries and to the slice header size, a set of parameters that has to be transmitted at the beginning of each slice. Because of this, new approaches aiming at facilitating parallel processing have been adopted in HEVC, described in the following two sections.

B. Wavefront Parallel Processing

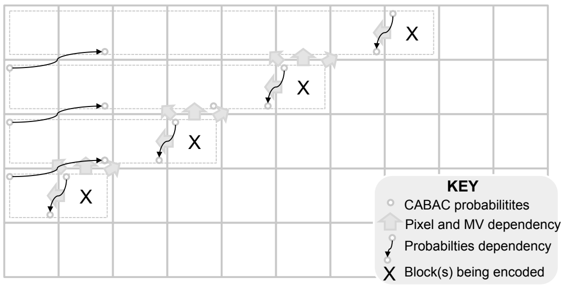

Figure 5: Wavefront Parallel Processing.

In [6], wavefront parallel processing (WPP) is proposed in order to enable parallel encoding and decoding (**Figure 5**). WPP consists of resetting the CABAC probabilities of the first LCU in each line with the probabilities obtained after processing the second LCU of the line above. Otherwise, all inter-block dependencies are maintained. Thus, parallel encoding and decoding is possible with moderate BD-rate degradation (around 1.0% compared to a non-parallel friendly bitstream in random access configuration [7]). Because dependencies are not broken, it is possible to convert a non-parallel compressed video to and from a parallel-friendly one (this is simply an entropy level operation). Furthermore, Cabac encoding is flushed after the last LCU of each row, making the bitstream representing each row of LCU accessible using entry point defined in the slice header. Thus, it is possible to use any number of core between one and the number of LCU rows in the frame in the decoder or in the encoder.

C. Tiles

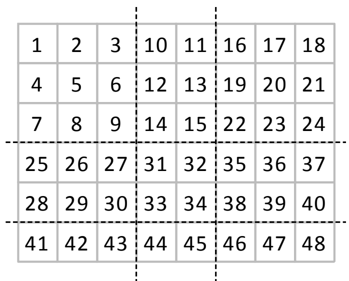

Figure 6: Frame partitioning into tiles.

In HEVC, a frame can be partitioned into a number of independent tiles [8] (**Figure 6**). Tiles are rectangular groups of LCUs. Tile boundaries are vertical and horizontal and extend across the whole picture. Tiles are processed in raster scan order, and the LCUs inside each tile are also processed in raster scan order. All dependencies are broken at tile boundaries, so there can be no pixel, motion vector or context prediction across them. The entropy coding engine is reset at the start of each tile. Only the deblocking filter is applied across tiles, in order to limited visual artifacts. Consequently, tiles can be encoded and decoded by independent cores working in parallel, and only the deblocking stage requires cross-tile communications. This comes at the expense of rate-distortion loss (about 2.7% loss compared to a non-parallel bitsream in random access configuration [7] when the same degree of parallelism as WPP is used). HEVC tiles are similar to JPEG2000 tiles [12] except that the tiles may have different sizes inside one frame, and they may have dependencies since deblocking can be performed across them.

D. Performance with tiles and wavefront tools

In [10] preliminary implementation of parallel HEVC decoding tools on a PC based platform give promising results that have been improved in [11]. In [11] authors have performed a parallel implementation of both proposed parallel tools on a 6-core dual processor at 3.33GHz. Without any further optimization to the reference software except the use of parallel primitives, [11] shows a real-time decoding on 12 cores for 1080p sequences. In TABLE I. one can notice that the acceleration on 6 cores is around 5 with the use of WPP or Tiles, adding 6 more cores lead us to only 10fps more. Due to its massively parallel nature the ALF part (20%) is almost independent from the number of threads [10].

TABLE I. WPP/TILES PARALLEL IMPLEMENTATION

Resolution	Cores	WPP (fps)	Tiles (fps)
1080p	1	8.8	8.7
1080p	6	40.2	42.9
1080p	12	50.9	57.6
2160p	1	1.5	1.5
2160p	6	8.1	8.5
2160p	12	13.2	14.8

IV. POTENTIAL APPLICATIONS

Regarding the compression gain already demonstrated, one cannot deny that HEVC will offer technical and commercial benefits to existing applications and usage scenarios. Hence, potential applications lie on IPTV (SD or HD) over DSL, where HEVC would increase service reach, or on point-to-point contribution on premium or light links. Moreover, multi-screen applications or OTT services can also benefit from HEVC by improving the overall quality of video to mobile devices.

On the other hand, HEVC enables future services not possible with today's state-of-the-art compression standards. Hence, potential applicable uses would be delivering 1080p60/50 at bitrates comparable to today's 1080i data rates, full resolution HD 3DTV at today's HD delivery rates. HEVC enabling better picture quality at lower bitrates, it will offer sports fans, for instance, a better viewing experience on mobile devices.

Last but not least, 4K applications for which HEVC was originally designed. Indeed, the broadcast industry is completing the transition to HD all around the world, and demand for an even enhanced end-user experience especially for home theater and premium events like sports is rising. End-users want to experience the event in an immersive way, as close as possible to the on-site viewing experience. 4KTV would be the answer and HEVC the straightforward video compression technology involved for 4K delivery to the home.

V. CONCLUSION

This paper gives an overview of the upcoming HEVC standard and details major changes or features in comparison with the well-established H.264/AVC standard. HEVC has been optimized for better-perceived visual quality on large image especially for Ultra High Definition (UHD at least Quad Full-HD or 4K). HEVC also provides tools to take advantage of multicore platforms in the sense that video bitstreams provide native hooks for this kind of parallelization with promising preliminary results (1600p at 30 fps) [11].

REFERENCES

[1] JCTVC-I1003, "High Efficiency Video Coding (HEVC) text specification draft 6", San Jose, Feb. 2012.

[2] B. Li, G. J. Sullivan, J. Xu, "Comparison of Compression Performance of HEVC Draft 6 with AVC High Profile," JCTVC-I0409, JCT-VC, Geneva, 2012.

[3] F. Kossentini, N. Mahdi, H. Guermazi, M. Horowitz, S. Xu, B. Li, G. J. Sullivan, J. Xu, "Informal Subjective Quality Comparison of Compression Performance of HEVC Working Draft 5 with AVC High Profile," JCTVC-H0562, JCT-VC, San Jose, Feb., 2012.

[4] TK Tan, A Fujibayashi, J Takiue,"[AHG8] Objective and subjective evaluation of HM5.0", JCTVC-H0116, JCT-VC, San Jose, Feb., 2012.

[5] M.Winken, P.Helle, D.Marpe, H.Schwartz and T.Wiegand, "Transform Coding in the HEVC Test Model," IEEE International Conference on Image Processing ICIP 2011, Brussels, 2011.

[6] C. Gordon, F. Henry, S. Pateux, "Wavefront Parallel Processing for HEVC Encoding and Decoding", JCT-VC document JCTVC-F274-v2, Torino, July, 2011.

[7] J.-M. Thiesse, J. Viéron, "On Tiles and Wavefront tools for parallelism", JCT-VC document JCTVC-I0198, Geneva, April, 2012.

[8] A. Fuldseth, M. Horowitz, S. Xu, A. Segall, M. Zhou, "Tiles", JCT-VC document JCTVC-F335-v2, Torino, July, 2011.

[9] J.R.Ohm, G.Sullivan, F.Bossen, T.Wiegand et al., "JCT-VC AHG report: HM subjective quality investigation (AHG22)," JCTVC-VC document JCTVC-I0022, San-Jose, Feb. 2012.

[10] M. Alvarez-Mesa, C. C. Chi, B. Juurlink, V. George, and T. Schierl, "Parallel Video Decoding In The Emerging HEVC Standard," Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), Kyoto, Japan, March 2012.

[11] C. C. Chi, M. Alvarez-Mesa, B. Juurlink, V. George, T. Schierl, "Parallel Scalability and Efficiency of WPP and Tiles," JCTVC-VC document JCTVC-I0520, Geneva, 2012.

[12] A. Skodras , Ch. Christopoulos , T. Ebrahimi, "The JPEG 2000 still image compression standard", IEEE Signal processing Magazine, Vol 18, pp. 36-58, 2001.