THOMSON
VIDEO NETWORKS

WHITE PAPER

Perspectives and Challenges for HEVC Encoding Solutions

Xavier DUCLOUX, December 2013

1. INTRODUCTION 3

2. HEVC STATUS 3

2.1 HEVC Standardization 3
2.2 HEVC Tool-Box 4

3. WHY SWITCH TO HEVC? 9

3.1 HEVC Performance 9
3.2 HEVC Complexity under Control and Real-Time Implementations already there 9

4. WHY THE EXPERTISE OF ENCODER PLAYERS IS KEY? 10

5. WHY HEVC IMPLEMENTATION CAME SO FAST? 12

6. FOR WHICH APPLICATIONS AND WHEN? 13

7. CONCLUSION 14

1. INTRODUCTION

The adoption of the new Joint ITU-T / MPEG video compression standard, named High Efficiency Video Coding, is a major event of the 2013 broadcast industry. Every time a new compression standard is released, the same questions are raised:

- What kind of video processing tools can explain the compression gain?
- Is the announced compression gain theoretical or really there?
- Can it be implemented for real-time applications?
- What deployment timeline and who will be the early adopters of the new standard?
- What commercial factors will lead to a deployment?

This White Paper will attempt to answer these pertinent questions.

2. HEVC STATUS

2.1 HEVC Standardization

HEVC, acronym for High Efficiency Video Coding, is a new video compression standard which has been developed by a Joint Collaborative Team of ISO/IEC MPEG and ITU-T VCEG (JCT-VC), as was the case for MPEG-2/H.262 and MPEG-4-AVC/H.264 in the past. After more than two years of collaborative work, the final specification text has been finalized and approved by ITU-T and ISO-IEC. A new International Standard was subsequently released.

The goal of the HEVC standard is to improve the compression efficiency to provide 50% bit rate reduction compared to the H.264/AVC standard, which has the best performance among the current generation of video compression standards and is widely used for High Definition Broadcast services.

Since the first meeting in April 2010, in which initial proposals were evaluated, there has been a growing interest around this new video compression standard, with more than 200 experts attending JCT-VC meetings and with a growing and large number of contributions evaluated.

The first version of the standard includes three video compression profiles, all dedicated to consumer applications:

- **Main profile**, which offers the best complexity/video quality compromise between the two initial profiles defined in the software reference model (Low Complexity and High Efficiency) and which works with 8-bit color depth samples,
- **Main 10 profile**, which has the same tools as the Main profile but supports 10-bit color depth samples,
- **Main Still picture profile**, which uses the same Intra tools as the Main profile and which works with 8-bit color depth samples.

The Main 10 profile has been adopted in the first version of the standard thanks to the initiative of a few broadcasters including DirecTV, BBC, BSkyB, NHK, SVT and equipment providers such as Technicolor and Thomson Video Networks [1]. This initiative aims at giving the choice to broadcasters to offer a better user experience for Ultra HD format with 10-bit color depth associated to a wider color gamut and avoiding any "legacy" issue if HEVC had been deployed with a single 8-bit depth profile.

Indeed, compared to the previous standard, the scope of JCT-VC studies has been enlarged to Ultra HD format (3840x2160 resolutions, commonly named as 4K). The Profiles of this first version are limited to 4:2:0 video contents, but Professional Profiles (4:2:2/4:4:4) with extra chrominance and bit depths are under study and will be released in April 2014.

The Joint Collaborative Team studies also multi-view and scalable extensions (SHVC) of the standard. These extensions will be published as amendments of HEVC in 2014.
New dimensions of the video signal, like wider color gamut and High Dynamic Range (HDR), are also under exploration for future amendments of the standard.

2.2 HEVC Tool-Box

HEVC is built on previous successful H.262/MPEG-2 or H.264/MPEG-4-AVC standards:

➔ **HEVC is still using a hybrid coding as in prior video compression standards:**

- prediction in the temporal domain followed by a suitable de-correlation technique in the spatial domain and scalar quantization (as illustrated in Figure 1)

➔ **HEVC shares many common elements with previous video compression standards (H.262/MPEG-2 and H.264/MPEG-4-AVC):**

- a conventional sampling of chrominance and association of luminance and chrominance data
- same picture types: Intra-predicted (I) pictures (pictures with Intra coded blocks only) or Inter-predicted (P or B) pictures (pictures with Inter-picture predicted blocks + possible Intra coded blocks)
- block motion displacement with variable block-sizes
- motion vectors over picture boundaries
- possible mix of Intra-picture and Inter-picture coding blocks
- block transforms (no wavelets or fractals)
- scalar quantization

[1] **"On a 10-bit consumer-oriented profile in HEVC", NGcodec, BSkyB, NHK, DirecTV, SVT, Motorola Mobility, Technicolor, Ericsson, Thomson Video Networks, BBC, ST - JCTVC-K109, Shanghai, Oct 2012**

➔ **HEVC shares many common elements with H.264/MPEG-4-AVC:**

- a hierarchical coding of pictures with multiple levels of B reference pictures
- spatial directional predictions for Intra coding
- multiple reference frames for Inter-picture coding
- motion compensation with ¼ pixel accuracy
- weighted predictions
- Integer Transform approximating a DCT
- In-loop adaptive deblocking filter
- same fine quantization scale
- Context-based Adaptive Binary Arithmetic Coding (CABAC)
- same encapsulation of video coding layer in a Network Adaptation Layer (NAL) as in H.264/MPEG-4-AVC
- same Sequence/Picture Parameter Set concept, to transmit In or Out-of-band information relevant to more than one picture

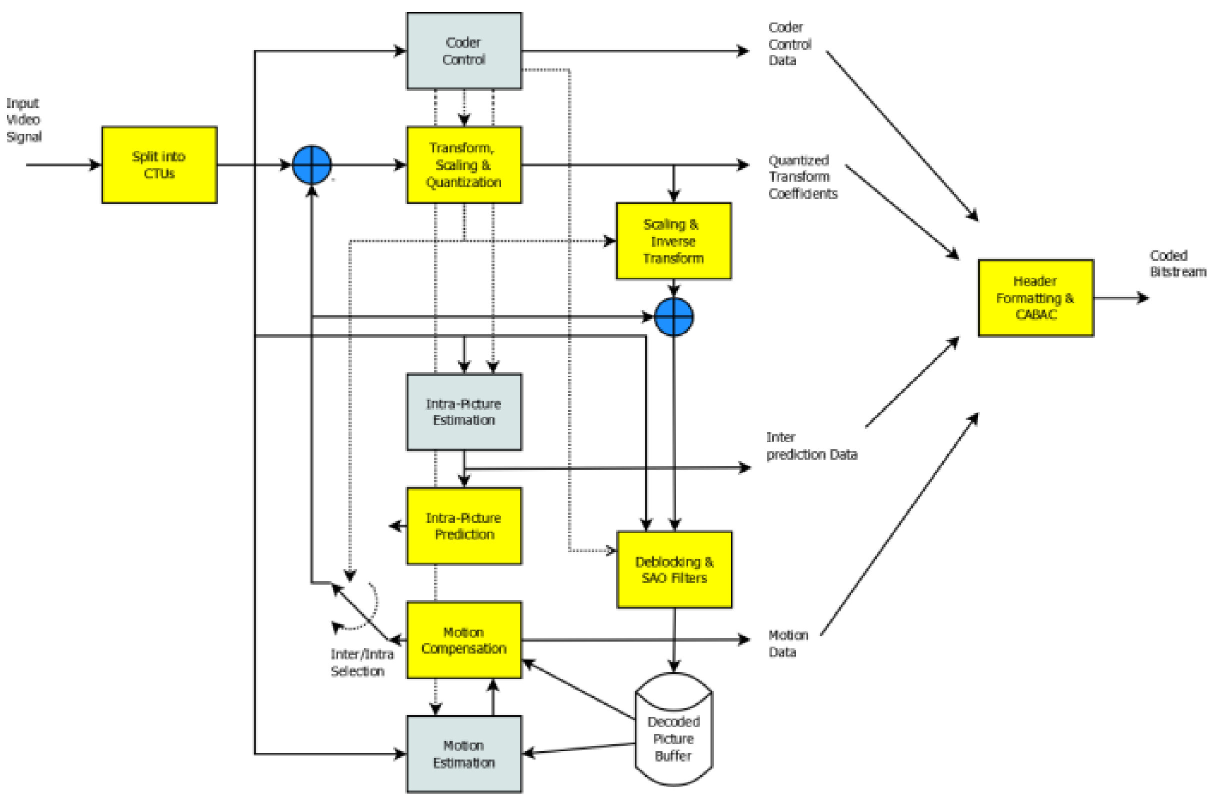

Figure 1 – Functional Blocks of a HEVC Encoder

But HEVC brings tools improvements to most areas:

➔ **a new recursive quad-tree approach** for the encoding structure gives a stronger adaptability and flexibility of predictions to the spatial and temporal characteristics of the video content (from very large 64x64 blocks to small 4x4 blocks) as shown in figures 2 and 3 below:

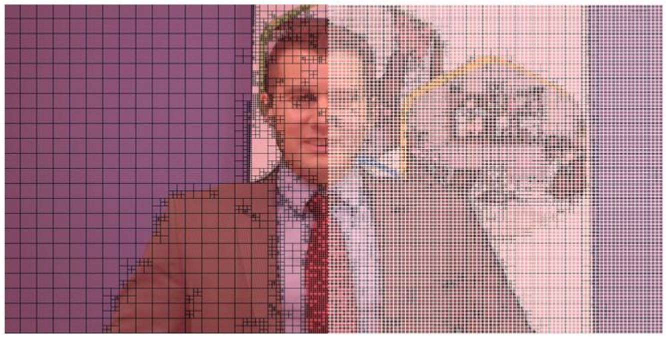

Figure 2 – HEVC Quad-Tree Coding Unit Partitioning (left) versus AVC Macroblock Partitioning (right)

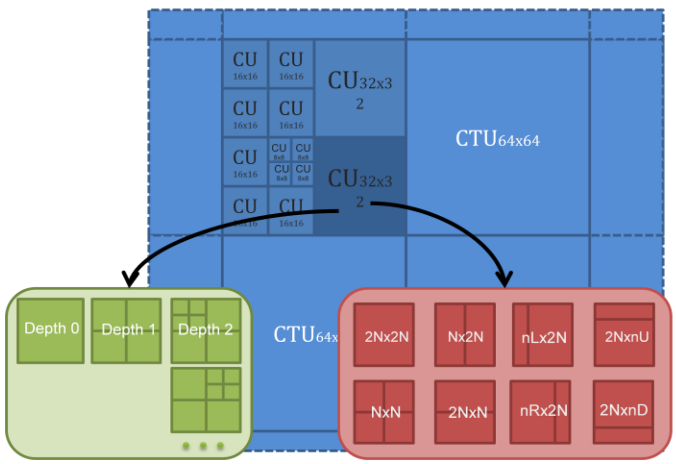

Figure 3 – HEVC Block (blue), Prediction (red) and Transform (green) Partitioning

➔ **improved Intra coding:**

- with an increasing number of directions for predictions (up to 35 instead of 9 in H.264/AVC)
- very large prediction blocks (up to 64x64 instead of 16x16 in H.264/AVC)
- Chroma predictions possibly inferred from Luma ones
- different scanning of Intra coefficients according to Transform size and direction

➔ **improved motion compensation:**

- with predictions units from 64x64 (16x16 in H.264/AVC) to 4x4 and possible asymmetric rectangular split of the block
- with the use of temporal and spatial correlation in motion vector prediction and for the construction of an implicit prediction mode, called merge mode (mode which limits prediction cost like direct mode in H.264/MPEG-4-AVC)

- with an improved 1/4 pixel and 1/8 pixel interpolation

➔ **extended Integer transforms:**

- with a Quad-Tree approach from the block partitioning
- may exceed the size of the prediction block
- can go up to 32x32, instead of 8x8 in H.264/AVC

➔ **the addition of an In-loop adaptive filter**, called Sample Adaptive Offset (SAO), after the deblocking filter to minimize the distortion

➔ **an optimized CABAC coding**

	AVC High Profile	HEVC Main profile
Picture partitioning	16x16 Macroblocks	**Coding Unit quadtree structure (64x64 to 8x8)**
	Predictions Units (16x16 to 4x4) – square or rectangle	**Prediction Units (64x64 to 4x4). Square, rectangle or asymmetric partitioning into 1 CU.**
	8x8 and 4x4 transforms	**Transform Units 3-level quadtree (32x32, 16x16, 8x8, 4x4) into 1 CU**
INTRA coding	8 angular predictions + DC	**33 angular prediction + DC + Planar**
INTER coding	Luma interpolation 6-tap + bi-linear to ¼ pel	**Luma interpolation 8-tap to ¼ pel**
	Chroma bi-linear interpolation	**Chroma interpolation 4-tap to 1/8 pel**
	Motion Vector Prediction using spatial median	**Advanced Motion Vector Prediction using spatio-temporal predictors (2 PU)**
	DIRECT mode (up and left MB)	**MERGE mode using spatio-temporal predictors (5 PU)**
Quantization	Scalar quantization + dead zone	**Scalar quantization + dead zone Rate Distortion Optimized Quantization (not in the standard)**
Entropy coding	CABAC or CAVLC	**CABAC**
Parallelism	-	**Wavefront Parallel Processing / Tiles / Dependant slices**
In-Loop Filering	In-Loop deblocking filter (4x4 boundaries)	**Light In-Loop deblocking filter, parallel friendly. Sample Adaptive Offset (SAO) filter**

Figure 4 – Tools Comparison: AVC High Profile versus HEVC Main Profile

In addition, HEVC brings **parallelism tools** to ease multi-threads software encoding and decoding implementation:

- tiles, which provide a simple synchronization mechanism but rather coarse level of granularity for parallelism (a few regions of the picture) thanks to independently-decodable rectangular regions of a picture with some shared header information,

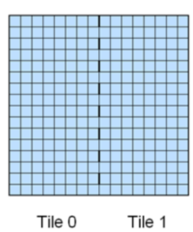

Figure 5 – Tiles Partitioning

- wavefront, which provides a more complex synchronization mechanism but rather fine level of granularity for parallelism (rows of blocks) thanks to CABAC context model inferred from the first two blocks of the previous row of blocks,

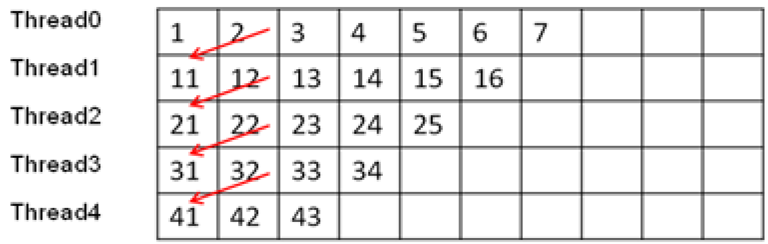

Figure 6 – Wavefront Mechanism

- dependent slices, which lay on wavefront mechanism but which carry data in separable NAL units.

From the beginning, HEVC standard has been **focused on progressive content**. Interlaced formats can be supported but are handled in a much simpler way than in previous standards:

- the two input fields can be encoded as separate fields in a pure field coding mode, i.e. fields pictures replacing frame progressive pictures, or in a pure frame coding mode, i.e. the two fields are interlaced together in one frame picture, but the choice is made for a complete sequence of pictures without any possible link between the sequences (closed-GOP constraint),
- no frame/field coding adaptation inside a picture.

An Ad-hoc group has been created in MPEG to evaluate the coding efficiency of HEVC for interlaced formats. Thomson Video Networks is actively participating to this group through:

- the characterization of interlaced content to make the selection of the reference set of sequences,

the delivery of HEVC reference model simulation results in comparison to state-of-the-art AVC encoder results for a subjective assessment.

3. WHY SWITCH TO HEVC?

3.1 HEVC Performance

Both objective (PSNR-based) and subjective video quality assessments have been performed within the JCT-VC and test results confirm that the initial targets can be reached:

- in its best encoding profile, using the full set of tools, HEVC can provide around 40% bit rate savings for equal PSNR for the 1080p sequences encoded in a long GOP structure (~1s I picture period, hierarchical group of 8 pictures),
- when considering equal subjective quality, bit rate savings are even greater: over 50% for all the test sequences and up to 70% for two of them,
- the performance of HEVC in pure Intra coding is also much better (~20% gain) than H.264/AVC or JPEG2000.

Subjective evaluations on various HD content show that encoding noise is significantly reduced compared to H.264/MPEG-4-AVC and that temporal stability is remarkable.

The low level of noise is due to the strong adaptation of prediction blocks and transforms to video content. The temporal stability can be explained by the improved motion coding, which is able to capture more redundancy with less signaling data.

The larger prediction units and transforms sizes give a better efficiency for the coding of homogeneous regions and the gain is (all the) greater as the picture resolution increases, as shown in Table 1.

Source type	Resolution	Frame rate	Bitrate saving average *	Bitrate saving min *	Bitrate saving max *
Progressive	3840x2160 (4K)	25	30.6%	22.0% (Ducks)	42.3% (Old town)
	1920x1080	50	29.2%	17% (Parkjoy)	46.3% (Old town)
	1280x720	50	24.7%	14.6% (Parkjoy)	36.6% (Old town)

Table 1 - Rate/PSNR Objective Results on SVT Set of Multi-Resolution Sequences
*** PSNR Bjøntegaard metric 1 HM7.0 MP / JM18.3**

3.2 HEVC Complexity under Control and Real-Time Implementations already there

Despite the increasing complexity of Intra and Inter prediction modes, the coding and decoding computation costs are very well mastered. Indeed, the encoding time of JCT-VC reference sequences with the HEVC reference software model is increased only by 30% and the decoding time by 60% in its best encoding profile and coding structure, compared to H.264/AVC reference software model. Though, the complexity ratio between the two standards is higher than these figures, because HEVC reference software model is written in a more efficient way than the H.264/AVC reference software model.

It is never easy to give a complexity ratio between two video compression standards, especially on the encoding side, but estimations derived from first real-time implementations give a ratio between 3 and 4 on a HD format. The progress made in the new generation of powerful multi-cores processors allows for this extra complexity within a reasonable density, making HEVC HD encoder economically viable in the short-term. During IBC 2013, HEVC 1080p real-time encoding was demonstrated on Thomson Video Networks ViBE VS7000 Video System.

On the decoder side, a few companies have demonstrated real-time performance with software implementations, even before the publication of HEVC standard, like:

- Samsung with HD real-time decoding on tablets,
- NTT Docomo with the following successive demonstrations:
  - a real-time 720p, 30fps decoding capability on a tablet (using a single core of an ARM Cortex A9 processor clocked at 1 GHz)
  - a real-time 1080p, 25fps decoding of a ~3Mb/s stream on a smart-phone (lower bitrates for 30 fps)
  - a real-time 4K, 60fps decoding on a laptop with quad-core i7 processor

Since then, many companies have provided real-time software decoding solutions on tablets, laptops and PCs.

Thomson Video Networks made joint demonstrations of Ultra HD, 50/60fps real-time decoding capability at NAB 2013 with NTT Docomo and with Fraunhofer HHI at IBC 2013, using Thomson Video Networks' ViBE VS7000 Video System for 4K file encoding and Sapphire server for 4K playout.

Some preliminary hardware decoding implementations have also been demonstrated:

- ETRI for example showed at the last JCT-VC meeting in January 2013 a hardware implementation of a HEVC decoder, achieving real-time for 1080p, 60fps while running at 266MHz. From their experience, a hardware implementation of a HEVC decoder requires twice the logic count and 50% more SRAM memory than H.264/AVC, which means that integration into state-of-art semi-conductor technology should not be an issue
- Broadcom announced at the CES 2013 the first Ultra HD video decoder solution based on a 28nm ARM-based chip

At IBC 2013, Thomson Video Networks showcased live MPEG-DASH/HEVC streaming and decoding on a Qualcomm® Snapdragon™ 800 processor.

First decoder chipsets able to support HD and Ultra HD at 25/30 fps are now available and the next generation able to support Ultra HD at 50/60 fps should appear in 2014.

4. WHY THE EXPERTISE OF ENCODER PLAYERS IS KEY?

Encoding decisions (blocks sizes, predictions sizes and types, transform sizes) are highly computational and the recursive quad-tree approach of HEVC has drastically increased computations.

The picture is partitioned into blocks (called Coding Tree Blocks) of a given size (typically 64x64 pixels)and handled in a raster scan format. The following decision scheme is applied for each CTB:

- the best prediction and the best transform partitioning (minimization of the rate distortion criterion) is selected
- the CTB is then split into 4 square sub-blocks
- for the first of these sub-blocks, the above selection process of the best prediction and transform partitioning is made again
- this sub-block is itself split into 4 square sub-blocks and the prediction and transform selection process is made again
- this Top-Down quad-tree decomposition is made down to the smallest possible block size as illustrated in Figure 7
- a Bottom-Up quad-tree re-composition is then made to take the final block partitioning decision (red bullets in Figure 7). For example, when the best prediction is chosen for the 16x16 block and the 4 best predictions are chosen for the 4 8x8 blocks which compose this 16x16 block, the final block partitioning is chosen between the 16x16 and the 4 8x8

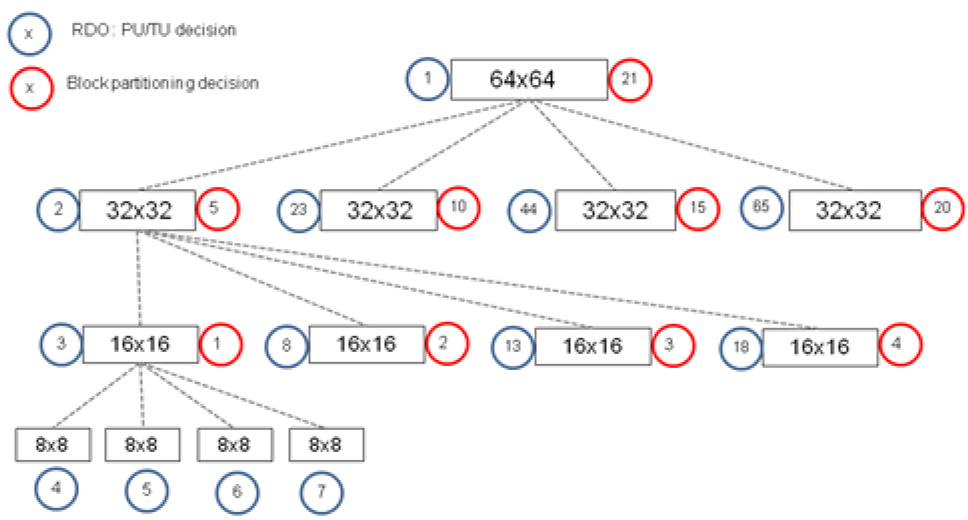

Figure 7 – RDO and block Partitioning Decision Tree

This recursive approach can be supported for small formats but can be rapidly expensive in terms of computation for the implementation of larger formats like Ultra HD. Alternative approaches can be developed to fulfill the two following goals:

- spare computation resources to increase density or reduce price of the encoding channel,
- maintain the perceptual quality compared to the “brute force” recursive approach or even more, increase this quality through more stable encoding choices, taking into account both spatial and temporal dimensions.

To that effect, some fast-decisions algorithms can be used to select the appropriate block partitions, block predictions sizes and transform sizes, based on pre-analysis of video content or on encoding context. At the heart of the block-level decisions, the local quantization of blocks, which is out of the scope of the standardization, plays a major role in the final rendering of video.

That is where the strong knowledge and expertise acquired with the previous standards is key.
On top of this block-level decisions process, it is worth mentioning that there are other parts of an encoder, where encoder-makers can bring some added value to the customers:

- motion estimation,
- picture type and ordering, picture structures, in function of sequence events (scene cuts, fades, interlaced/progressive scan),
- rate control and statistical multiplexing expertise.

Mastering these functions remain essential to achieve the best and the most stable quality.

5. WHY HEVC IMPLEMENTATION CAME SO FAST?

Even if the block-level computation complexity is significantly increased, there were some arguments for saying that real-time implementation of HEVC encoding with good performance would go faster than for AVC/H.264:

- in AVC/H.264, CABAC entropy coding was a new and very complex tool compared to previous simple entropy coding methods used in MPEG-2 and had been a real bottleneck for a fast deployment of AVC/H.264. The CABAC entropy coding has been reused in HEVC and mainly optimized compared to AVC/H.264. Implementation is therefore straightforward,
- AVC/H.264 has also introduced a new concept in the handling of group of pictures (GOP) , using a hierarchical GOP coding structure with the use of multiple-level B reference pictures and more complex reordering of pictures than in MPEG-2. Moreover, this new and very powerful concept did not come with the first version of AVC/H.264, but later, during the scalable extensions of AVC/H.264, called SVC. This new tool for GOP handling adds some complexity in the picture decision algorithms with P periods, being adapted dynamically from 1 to 8. This hierarchical GOP coding structure is reused in HEVC, which means that multi-years fine-tuned picture decision algorithms can be reused straightforward,
- the encapsulation layer (NAL) of the HEVC video stream remains the same as AVC/H.264, which means that the HEVC stream can be mapped into the transport layers used in the broadcast transmissions (MPEG-2-TS) or streaming (ISOBMFF or TS File formats) with very small work (new NAL types to consider),
- A part from these common tools shared with AVC/H.264, HEVC offers also parallelization tools, which can ease software implementation on multi-core processors without degrading compression performance. This is of course a strong argument since multi-core processors are massively used now. These tools will allow real-time implementation of Ultra HD encoding in a rather short term.

6. FOR WHICH APPLICATIONS AND WHEN?

The adoption of HEVC for mobile services on tablets or smart-phones should be very fast, because of:

- the fast renewal of decoding devices (two years in average) in this type of application, making a change of standard possible,
- HEVC HD real-time software implementation already demonstrated on mobile devices,
- the ability to offer HD services over cellular networks, thanks to the bit rate performance improvement over H.264/AVC, going together with the increasing resolution of the mobile devices.

The deployment of HEVC for mobile services (VOD, live) could start from 2014. For a deeper analysis, please refer to the complementary White Paper "HEVC, MPEG-DASH and eMBMS: three enablers for enriched video contents delivery to handheld devices over 4G LTE networks" on Thomson Video Networks' website.

News Gathering and high-end contribution applications such as sporting events could also take benefit of lower bandwidth required by HEVC and avoid the constraints of a big existing decoder park. They could be a fast adopter in a one or two year timeframe.

For IPTV applications, the lower bandwidth required by HEVC will extend subscribers eligibility to HD services. First decoder chips are now available and many cheap set-top-boxes (STBs), which are necessary for the deployment of such applications, should be demonstrated at the CES event in January 2014. . There will not be a full replacement of existing H.264/AVC STBs, since many IPTV operators have recently launched a new generation of these STBs, but instead, the deployment of HEVC STBs to new customers, which would be more cost-effective than pushing fiber closer to homes. On OTT side, HEVC-based initiatives for on demand services were announced for this year, together with first HEVC-ready connected TVs announcement at the CES event in January 2013,but it seems that the real take-off of HEVC on demand services will be in 2014. Netflix and Sony Picture made some announcements for the delivery of premium Ultra HD services from 2014.

Cable applications do not have the same bandwidth issue as IPTV or OTT services. That is why MPEG-2 compression is still used in many countries and the massive existing set-top-box park does not ease the migration to a new compression standard. Though, Ultra HD, bringing more immersive content, could be an interesting differentiator in the cable operators offer compared to other media and could be a strong incentive for the deployment of new STBs from 2016. HEVC is clearly the right answer for this Ultra HD deployment.

For Digital Terrestrial Television applications where decoders are integrated into the displays, the replacement of existing TV sets is clearly a legacy issue. In addition to HEVC decoding issue, the question of the migration to DVB-T2 modulation is raised for all countries which are still using DVB-T. A few European countries have recently taken some initiatives to put into the law, as soon as possible, the obligation for receivers to support both DVB-T2 and HEVC standards, in order to be able to launch services from 2015-2016. First experimental Ultra HD channels could also be launched at that time. In France, the TV watchdog, CSA (Conseil Supérieur de l'Audiovisuel), considers that an Ultra-High-Definition Television (UHDTV) channel could be launched in 2018.

For Direct-To-Home applications, the question of the replacement of large existing decoder park is clearly an issue. But as in cable applications, Ultra HD could be seen as a strong and interesting differentiator for satellite broadcasters and HEVC is the appropriate answer to deploy this new format. With the help of Thomson Video Networks, Hispasat launched a permanent Ultra HD AVC + HEVC channel over Europe, named Hispasat 4K, during IBC2 013, to promote the adoption of Ultra HD and facilitate interoperability tests.First commercial Ultra HD services over satellite have been announced for 2014 in Japan and should appear the year after in Europe or in the US.

7. CONCLUSION

Video consumption over any kind of media (satellite, terrestrial, cable, Internet, mobile networks) will continue to grow to reach unprecedented levels over the next few years. The new joint ITU-T / MPEG standard, HEVC, is bringing an interesting bit rate gain of 50% versus H.264/AVC, which makes it the right candidate for a massive deployment in the coming years:

- either by offering a new immersive experience with Ultra HD
- or by providing capability to offer HD services to more and more consumers on any kind of media
- or by solving bandwidth bottleneck on media which face an ever-increasing video demand, like Internet or mobile networks.

CONTACT INFORMATION
Please contact your sales representative or visit our website for all contact details. http://www.thomson-networks.com