High Efficiency Video Coding (HEVC) in a Changing World—What Can MSOs Expect?

Mukta Kar, CableLabs; Yasser Syed, Comcast Cable; Munsi Haque, Consultant

Abstract

The cable operator world has been undergoing a sea-change over the last couple of years. Content is increasingly being viewed in a non-linear fashion. Service providers are not just delivering to the leased set-top box (STB), but to PCs, gaming machines, tablets, cell phones and other customer owned and managed (COAM) devices. The migration of Picture quality is not just from standard definition (SD) to high definition (HD), but also to 3D, 4K (also called UltraHD), and several resolutions in-between. STB Video Processors are changing from dedicated hardware processors to general-purpose multicore processors running video processing applications.

Amidst these changes, MPEG High Efficiency Video Codec (HEVC, also called H.265), the new video coding standard released by ISO/IEC & ITU-T has just been released in its version 1 format in January 2013. It brings an additional 2:1 compression efficiency over it predecessor Advanced Video Codec (AVC) and incorporates several improvements suitable for video deployments in this new environment. This paper will examine the new improvements of HEVC, including compression performance, and what areas it may be employed to enhance in this new and changing service operator environment. Lastly this paper will conclude with integration/migration strategies to introduce HEVC technologies into Cable services, including TV Everywhere services.

INTRODUCTION

Video compression technology coupled with MPEG standardization brought a new era in video delivery and applications. The deployment of MPEG-2 video in broadcast video has been a huge success in terms of improved video quality and increased number of channels. It also increased consumer choice in the area of video services due to emergence of direct broadcast satellite (DBS), DVD and IP delivery. The MPEG-2 standard is broadcast centric and not friendly to other video applications, especially to real-time internet delivery. As the internet is a public network, and is based on a best efforts delivery protocol, streaming video delivery over the internet lacks in video quality and resolution compared to broadcast video. But, there are two primary advantages of HTTP internet delivery over broadcast delivery: 1) video/audio content can be delivered to any receiving device with internet connectivity and 2) content can be delivered in a personalized manner. To help reduce the bandwidth needed for video delivery and to address a wider area of video applications, the Joint Video Team (JVT) of MPEG and ITU-T published AVC/H.264 video compression standard in 2003 [9]. AVC provides 2:1 compression gain over MPEG-2 video. This acted as a catalyst for explosive growth in video applications, especially video over the internet. Although the internet was generally developed as a non-real-time data delivery network, numerous video applications, real-time and non-real-time, are now using nearly 50% of the internet bandwidth capacity.

Obviously this is impacting other services delivered over the internet due to real time bandwidth delivery demands. To mitigate the negative impact from video delivery over the internet, MPEG and ITU-T formed a Joint Collaboration Team for Video Coding (JCT-VC) and initiated another compression standard known as HEVC/H.265 in 2010 which provides significantly better compression than that of AVC. After working nearly 2-1/2 years, JCT-VC finalized HEVC version 1 standard in January 2013 [1, 2]. Again HEVC provides nearly 2:1 compression gain over AVC.

HEVC: DIFFERENCES FROM AVC

The development of HEVC started approximately a decade after AVC was started, but essentially is still an evolution of AVC with enhancement and refinement to some AVC tools, and with the addition of a few new tools (See Figure 1).

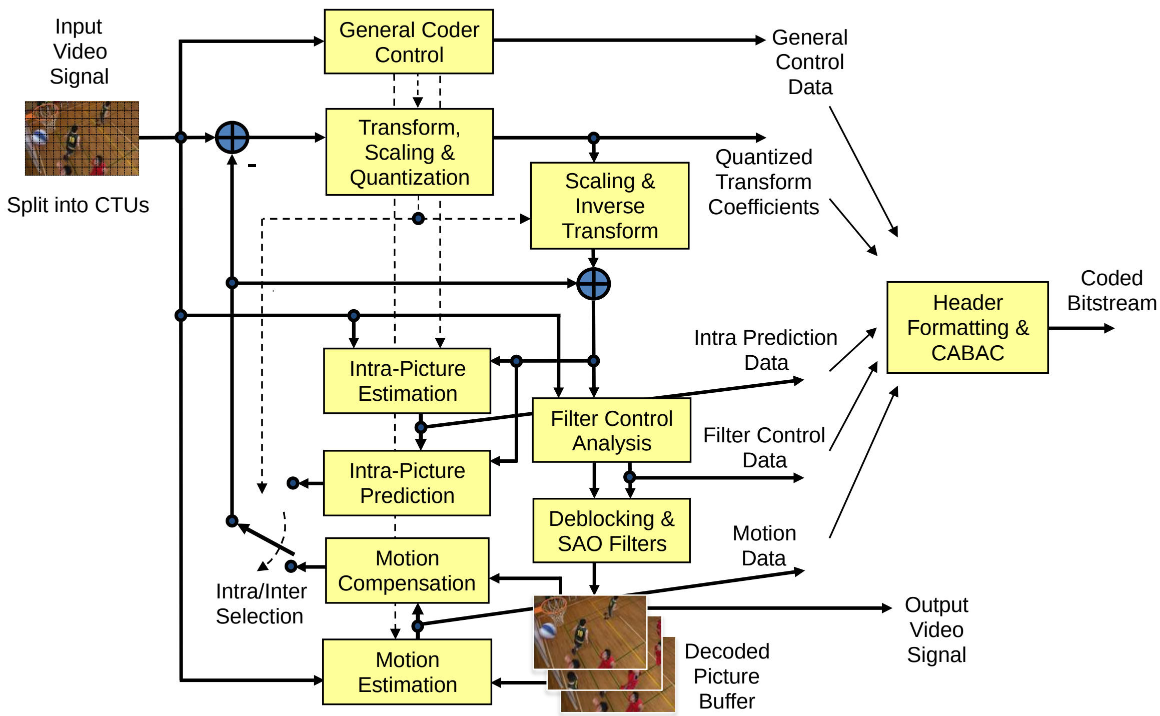

Figure 1: Typical MPEG HEVC Video Encoder Structure [10]

The new HEVC compression tools can be categorized around three main areas while maintaining the same or better visual quality by providing: 1) improvements to reduce number of bits required for region representations (Coding Unit, Transform Unit), 2) improvements for better prediction accuracy and reduction of errored residuals (e.g., Prediction Units, Spatial directional Modes, Adaptive Quantization), and 3) improvements in informational compaction/ symbol rates in bitstreams (simplified CABAC, New Scanning Modes). Table 1 below describes some of the evolution in common encoding tools from MPEG-2 to AVC to HEVC.

Coding Tools	MPEG-2	AVC	HEVC
Intra-prediction	None	Yes (9 modes)	Yes (35 predictions)
Inter-prediction	Yes (No B-picture as reference)	Yes (allows hierarchical b-picture as reference)	Same as AVC
CU (coding unit) size	16x16 (fixed, known as Macroblock (MB))	16x16 MB (same as in MPEG-2 video)	Variable, 64x64, 32x32, 16x16, and 8x8
PU (prediction unit) size	16x16	16x16, 16x8, 8x16	32x32, 16x16, 16x8, 8x16 8x8, 8x4, 4x8, 4x4
TU (transform unit size)	8x8 (DCT floating point)	8x8 and 4x4 (DCT integer)	32x32, 16x16, 8x8, 4x4 (DCT integer and also 4x4 DST integer)
In-loop filter	None	One Deblocking filter	Two in-loop filters (deblocking and SAO)
Entropy	VLC	CAVLC and CABAC	CABAC only
Parallel Processing tool	None	None	Tile and Wavefront

Table 1: Evolution in Common Video Encoding Tools

It is to be noted that the coding unit (CU) is analogous to the AVC macroblock but in this case the macroblock can change in size (see Figure 2) which when used appropriately can lead to bitrate savings. (See Table 2 [8].) Similarly, compression algorithms adaptively determine the size of prediction unit (PU) and transformation unit (TU) to achieve savings in bits while the picture quality is maintained at a desired level.

To aid in quality improvements, AVC uses a Deblocking Filter as an In-loop filter; HEVC uses a simpler but comparable Deblocking Filter and also adds a new Sample Adaptive Offset (SAO) Filter as the In-loop Filters. For Inter-picture prediction, HEVC uses Quarter-sample precision for the motion-vectors, and 7-tap or 8-tap filters for interpolation of fractional-sample positions. Whereas, AVC uses 6-tap filtering of half-sample positions followed by linear interpolation for quarter-sample positions.

HEVC also has three new features (Tiles, Wavefront Parallel Processing, and Dependent Slice Segments) to enhance parallel processing capabilities or modify slice data structures for packetization purposes. Such features help in an encoder or decoder implementation to derive benefits in particular application contexts.

Lastly HEVC also provides enhanced High Level Syntax bitstream support to improve operations over a variety of applications, network environments and robustness to data losses.

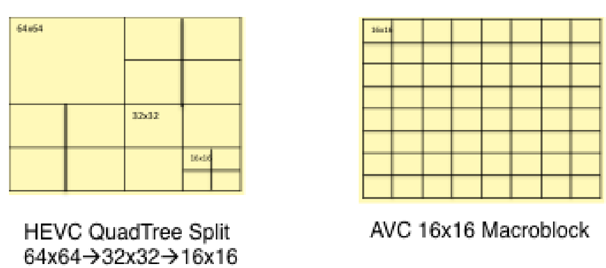

Figure 2: MPEG HEVC Coding Unit (CU) compared to MPEG AVC Macroblock

	Entertainment Applications		Interactive Applications	
	Maximum CU Size		Maximum CU Size	
	32 × 32	16 × 16	32 × 32	16 × 16
Class A	5.7%	28.2%	–	–
Class B	3.7%	18.4%	4.0%	19.2%
Class C	1.8%	8.5%	2.5%	10.3%
Class D	0.8%	4.2%	1.3%	5.7%
Class E	–	–	7.9%	39.2%
Overall	2.2%	11.0%	3.7%	17.4%
Enc. Time	82%	58%	83%	58%
Dec. Time	111%	160%	113%	161%

Table 2: Example of bitrate increase from 64x64 CU size to 32x32/16x16 CU [7]

In brief, enhancements/refinements have been done to some AVC tools with addition of a few new tools as well. The effect of these improvements has been the ability to use existing AVC tools in combinations to more precisely allocate bits in alignment with visual perceptual models as indicated by Figure 3 [6]. This allows for future advances in compression efficiency as image analysis algorithms improve. In addition, the reduction in bitrate and new parallel processing tools align well with battery-enabled portable devices. This is done through the use of low-power multi-processors CPUs rather than a single high-power processor and helps in new models of how video is consumed (connectedness, conferencing, portability, and on-demand).

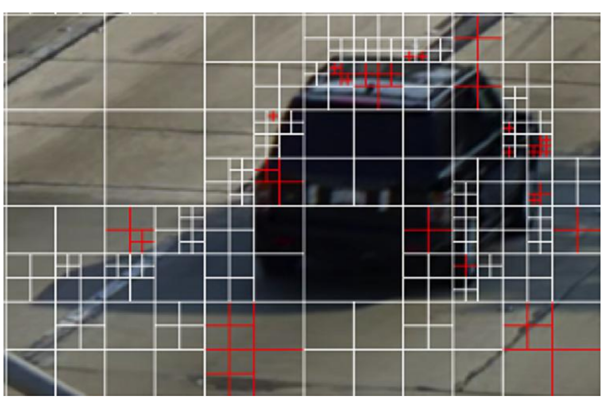

Figure 3: Example of recursive quad-tree partitioning for coding block (white) and transform block (red) [3]

PROFILES AND LEVELS

Profiles, tiers and levels specify conformance points for implementing the standard in an interoperable way. A profile defines a set of coding tools or algorithms that can be used in generating a conforming bitstream, whereas a level places constraints on certain key parameters of the bitstream such as maximum picture size, maximum bit rate, and a few other parameters which basically relate to decoder processing load and memory capabilities. In the design of HEVC, it was determined that two distinct sets of applications exist that have requirements that differ only in terms of maximum bit rate and CPB capacities. To resolve this issue, two tiers were specified for some levels – a “Main” tier is intended for most consumer video applications and a “High” tier for higher quality delivery that requires much higher bitrates. A level using the “Main” tier only needs to encompass levels below with bitrates targeted for the “Main” tier at that level. The decoders conforming to a specific profile must support all features in that profile.

Currently, three profiles, called the “Main”, “Main 10” and “Main Still” have been specified. Main is intended to be used with video with pixel depth of 8 bits and Main 10 is for 10 bit video. Main and Main 10 have the same set of tools, and a Main 10 decoder can decode a Main 8 compliant bitstream, however, the reverse is not true. It is possible that some other profiles of the standard will also be specified. Minimizing the number of profiles provides a maximum amount of interoperability among devices, and across applications such as broadcast, mobile, conferencing and streaming. Thus, a Main profile compliant device may be used for more than one application. Main Still is intended to be used for still picture decoding.

The objective here is that devices like cameras, smart phones and other similar devices can capture or decode both video as well as still pictures without the need to support two different codecs as is the case today.

The definition of 13 picture levels has been defined, starting with picture sizes such as a luma picture size of 176×144 (sometimes called quarter common intermediate format) to picture sizes as large as 7680×4320 (often called 8k). There are two tiers supported for eight of these levels (levels 4 and higher); Main tier and High tier. A decoder supporting High tier will be able to decode Main tier compliant streams, while the reverse is not true. High tier decoders support bitstream with higher maximum bitrate and hence may need higher processing power. In most cases high bitrate bitstreams provide higher video quality than the one provided by a lower bitrate bitstream.

PERFORMANCE IMPROVEMENTS

It is noted earlier that HEVC provides 2:1 better compression gain over AVC. So to deliver content with same quality, a HEVC codec will use approximate half the bitrate of that required by AVC. In other words with bitrate close to the AVC bitrate, a HEVC coded video can provide significantly better video quality. Again the compression gain can vary with content and also picture sizes. Larger size pictures tend to provide better compression efficiency than smaller resolution pictures. The table below shows the typical bitrates needed by MPEG-2, AVC and HEVC bitstreams. In subjective tests results with equivalent reproduction qualities for codecs, Table 3 shows HEVC encoders use approximately 50% less bit rate on average than AVC encoders. HEVC design is especially effective for low bit rates, high-resolution video content, and low-delay communication applications.

Resolution/FPS	HEVC	AVC	MPEG-2
UHDTV- 4K/60	8-15 Mbps	18-22 Mbps	High
HD- 1080/720P60	1.5-3.5 Mbps	5-9 Mbps	9-15 Mbps
HD- 720p30	0.8-2.0 Mbps	1.5-4 Mbps	3-5 Mbps
SD	0.4-0.7 Mbps	0.7-1.5 Mbps	2-3 Mbps

Table 3: MPEG HEVC Performance over Typical and Anticipated Video Services (Linear/VOD)

Performance Stats	HEVC over AVC [High Profile]
Encoder Complexity	~10x (or less)/Early 4x
Decoder Complexity	~1.4x
Memory	~1.25x
Memory Bandwidth	~1.25x

Table 4: MPEG HEVC Anticipated Complexity, Memory, and Memory Bandwidth Performance

Real-time software decoding [3, 4, 5] of HEVC bitstreams is very feasible on current generation devices — 1080p60 decoding on laptops or desktops, and 480p30 decoding on mobile devices.

As an example-1, HEVC software decoding of 480p to 1080p at 25 or 30 fps is possible with a single core of an ARM processor. Here the player application itself can be multi-threaded, with separate decoding and display threads (OpenGL to display video in real time). Also scaling to fit screen size and YUV to RGB color conversion can be done on the GPU during shading.

As an example-2, the software playback of 4K sequences at 60 fps is possible on a laptop where bitstreams encoded with random access main profile at 12 Mbps bit-rate can be decoded by using three parallel decoding threads with a quad core, 2.7 GHz, Core-i7 processor. Up to 100 fps are achieved with four parallel decoding threads on the same laptop.

MPEG HEVC encoders are expected to be a few to several times more complex than AVC encoders, and a subject of research in years to come. The market availability of real-time 1080@60P encoders supporting HEVC Main/Main 10 profile is expected to be available by 2014.

Service Providers: WHAT's ALREADY CHANGING?

Over the last few years, the types of services being supported have significantly evolved (see Figure 4). Previously customers watched video services through a cable set-top box and on one of the 2-3 TVs in the house. Data connections were meant for the PC and laptops available in the house. With the advent of wireless systems, those PCs, laptops, and now phone devices became portable, but bandwidth reliability was still causing interruptions and used mainly for data services rather than video viewing. With higher performing data devices, and the advent of MPEG-AVC, and settling some of the firewall issues, these devices are starting to favor video viewing more often. At the same time, television screens are getting bigger, HD is more firmly entrenched in households, and various flavors of non-linear viewing (on-demand / DVR) are increasing in popularity. With these establishing trends, the service expectation is evolving to view video on a plethora of devices (cable box, gaming console, Blu-Ray Player, Boxee Box, Roku, Tivo, Tablet, and Smartphone) and not just something that is permanently fixed in the living room. The choices made by the customer then are determined by the perceived value of the service.

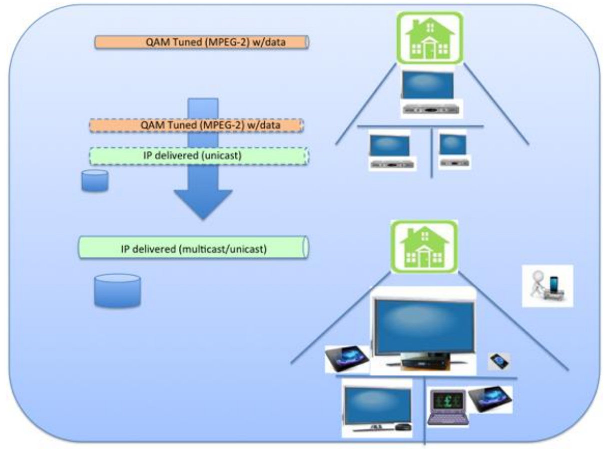

Figure 4: Household Past and Present/ Future Video Device Consumption Models

So what makes content more valued to the customer? Is it the material? Partially. But it also has a lot to do with connection and experience as well (see Figure 5). In terms of content, the value can be affected by how it can be purchased (subscription, single time, catalog depth), who is excluded from getting it, and by when it can be available (first to air, first available). In terms of experience, value can be affected by how immersive it is (large screen/4k/8k/portable screen), where it can be watched (fixed, mobile), how it can be shared (group/individual setting), and lastly how it can be consumed (languages, CC, Binge viewing). There are some exciting things happening in the area of experience. The content can be more valued because it is larger (more resolution), faster (more frame rate), and brighter (more contrast). Personal/Portable devices (PC/tablets) in the home are rapidly replacing the 3rd or 4th TV set in the household. The value of content within the household increases in this mode due to added accessibility of the content and transferability of it (your child can now watch anime on his/her laptop). The value further increases if one can then walk out the door and still can watch the content. The tradeoff on this is what the risk is to secure the content versus mobility, and a compromise to this would perhaps be a download-to-go application. In addition to increased resolution there is another dimension of increased frame-rates (60 Fps, 30 Fps and 3:2 pull-down). Lastly with advancements in cameras, images can be captured with higher contrast due to improvements in sensors which leads to shooting of a lot of dark, high-contrast scenes. In this case the experience has expanded, but is only allowed if the implemented technology can support it.

The last component is connection (which is more hidden to the service) that deals with more infrastructure and technology aspects. The value here can be affected by the availability of the connection (always-on, managed, and unmanaged), the distribution of the content (broadcast, unicast, DRM, and link encryption), the host processor capabilities (STBs, tablets, software based, hardware based), and lastly by request (linear, non-linear). Note the advent of IP delivery has opened up a set of new customer owned and managed devices. The choices of using broadcast or IP unicast can limit or expand the audience size when demand requires it. In earlier periods, the infrastructure design deemed all content valuable and protected the content all equally, but restricted content to the confines of the infrastructure. With the advent of DRM-protected content, the infrastructure restrictions were loosened, all content did not have to be protected equally, and it became more mobile. Lastly technical capabilities, like codec format, can affect capacity of the connection because the codec can affect things like the management of bandwidth, the capacity of storage, and the processing on devices. The calculations are for the value of content changes to include platform portability as well as increasing the weighting of accessibility and availability. In the real-time linear domain, this can increase the value of sports and news. While in the VOD domain, this can help determine if a cable subscription of aggregate VOD catalog is ordered, or whether an over-the-top (OTT) service account is enough. What is also interesting is that the infrastructure and technology can change but the service to the customer could remain the same.

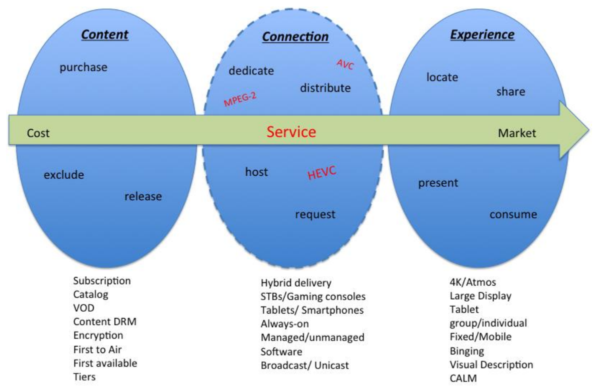

Figure 5: Content & Service Value Diagram

The service evolution that has been happening has affected the connection value aspects the most. With the new devices and new distribution pathways to get content in front of the viewer, there has been a subsequent 10x-100x increase of the number of streams and files to support new devices and unmanaged networks in the service operator ecosystem. These changes have put increased pressure on bandwidth and storage demands in the network due to support of adaptive streaming, expanded on-demand catalogs, and PVR services. Until just recently a single transcoder would have a single output stream. Presently a single input could generate up to 10 streams or files at the output. New distribution pathways such as on-demand MBR (Multi-BitRate) streaming for http adaptive streaming technologies has also increased bandwidth and storage demands to support bandwidth OTT congestion management and network storage.

The delivery of video is not only to a STB using an MPEG-2 transport stream over QAM, but now also to connected personal devices through an HTTP protocol over IP. For individual viewing of real-time content, the IP connection is very suitable to this medium. But when it comes to high valued content that is driven real-time high viewer demanded, often the MPEG-TS system has benefits that are sometimes more complicated to replicate in the IP domain. To repeat this experience in the IP domain would be an analogous multicast equivalent similar to the broadcast medium, but in the near term it is still replicated using a unicast IP delivery. Presently the storage may be commonly shared (used in HTTP adaptive streaming) but the delivery is still a unicast delivery which works unless one scales up in real-time viewership (e.g., the Super Bowl).

The service expectation is already changing to support better displays, more mobile and personal devices, and increased non-linear viewing. The service expectation has been moving in this direction because of the increase in value to the customer experience in the living room, and on the go. A large amount of improvement and value to the services has already happened with more expected to come. There will be increased pressure on the infrastructure due to higher pixel demands, increased number of output streams, and related storage requirements. Can our current technologies (Bandwidth, Storage, IP distribution) support the ramp up of these services as popularity grows? HEVC can be a good candidate to relieve some of the infrastructure demands as the service continues to grow.

POSSIBLE MPEG HEVC BENEFITS

HEVC can reduce the complexity and costs of handling multiple streams in this transitioning environment. For OTT-based services, the number of stream representations can be reduced since higher quality streams at lower bitrates can survive more often through bandwidth-congested environments. For a present HD OTT service, the number of streams could be reduced by 50% and in addition each of the remaining streams has also a reduced bitrate. For backhaul, bandwidth distribution demands for mezzanine/contribution streams can be reduced by switching from AVC/MPEG-2 formats to HEVC (or the quality of distributed content can be increased for the same bandwidth costs). This also encourages the transition from satellite-based feeds to fiber-based IP connections, even further reducing the signal integrity risks encountered with satellite distribution and redundancy strategies associated with it. In the area of non-linear services (VOD/Cloud DVR) and targeted ad insertion services, storage demands can be reduced by storing in an HEVC format even though output may be transcoded to a more traditional format. This becomes especially needed in light of supporting unique copy services across cable devices and customer owned and managed (COAM) devices. The reduction of storage demands can be greater than 50% due the decrease in number of streams supported and the decrease in bitrate to support the same quality streams. At the granularity of the stream for VOD, efficiency can occur by switching from using multiple trick files to support 2-3 speeds to supporting dynamic trickplay in the stream itself through the use of picture order count (poc) / temporal IDs in HEVC.

In terms of the device perspective, HEVC can increase the value associated with the host device by decreasing the service bandwidth to each device and increasing the amount of bandwidth available for the service. This benefit will assist in terms of IP unicast replications of expected broadcast customer experiences (e.g., popular events like the Super Bowl) while longer term solutions can then be created (e.g., multicast IP). Additionally HEVC does not require a new set of mobile personal devices to be created; existing mobile devices can already play 720p30 HEVC using software-based decoders taking advantage of the small increase in decoder complexity. Since mobile devices often have short (18 month) lifetimes, the transition to HEVC in this area can be easier with later introduction of higher valued content experiences. 720p30 HD streams are already the currently accepted deployment of HD streams for OTT services and smartphone/tablets devices. Lastly, battery lifetimes on the device (often consumed quickly by video viewing) can increase (HEVC streams with 4-8 hours of continuous play) by switching to an HEVC coding standard designed to take advantage of low-power multi-core processors in these devices. Additionally the reduced bitrate can save on antennae power and processing power required to decode on the portable device. This can also further increase battery lifetime.

In the areas with more immersive experiences and larger displays such as in the living room/bedroom, HEVC can greatly help out in this area by reducing the bitrate associated with carrying a higher quality or larger resolution stream. In terms of integrating these new experience streams, HEVC can reduce the bit rate to fit into the same slot as a present day MPEG-2 HD channel (~8-15 Mbps). With higher resolution streams, the gains could increase non-linearly (compared to HD) due to use of larger macroblock sizes, and increased spatial mode. With increased frame rates, HEVC can also reduce bit rate from adding additional frames using simplification methods/increasing accuracy in motion vectors and use of different options in prediction units. Hardware based decoders through silicon based chipsets are coming out early next year with 4K decoding that will add value to the larger display based devices that can be implemented through an ROI based approach based on 4K and higher services. Bandwidth demands on the plant can be managed with expectations set for new higher quality deployments (4K@60fps, 10 bit displays, etc.) within the same bandwidth of a traditional HD and in combination with other bandwidth saving strategies. The install base on this would coincide with an increased value in services that can be enabled on next generation STBs and gaming machines.

CONCLUSIONS

Cable services have evolved rapidly from a few years ago. Video services are now being offered for not just the living room, but also on the go for laptops, tablets, and mobile devices. An IP distribution structure using adaptive streaming techniques is used for these newer devices which enables portability in the managed home environment, WIFI, and OTT) delivery. Additionally there is an increase in number of streams delivered and managed as well as storage in cloud DVR and CDN structures. With the increased popularity of these services, there will be increasing pressure on the current infrastructure to support these services. Higher Efficient Video Coding (HEVC) can play an important role in deploying these new services on a larger scale by reducing the number and size of streams delivered and stored, making the IP unicast model more efficient for replication of older broadcast services, increasing the battery lifetimes on devices, and creating more immersive experiences for larger displays with higher frame rates and resolutions. HEVC can increase value by making the connection more efficient and improving the experience which can support the expansion of the service market.

ENDNOTES

[1] B. Bross, et. al., "High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Last Call)", JCTVC-L1003_v34, Geneva, January 2013

[2] G. J. Sullivan, 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, December, 2012.

[3] F. Bossen, B. Bross, K. Suhring and D. Flynn, "HEVC Complexity and Implementation Analysis", IEEE Transactions on Circuits and Systems for Video Technology, December, 2012.

[4] T.K. Tan, Y. Suzuki and F. Bossen, "On software complexity: decoding 4K60p content on a laptop", JCTVC-L0098, Geneva, CH, 14–23 January 2013.

[5] K. Veera, R. Ganguly, B. Zhou, N. Kamath, S. Chowdary, J. Du, I.-S. Chong, and M. Coban, "A real-time ARM HEVC decoder implementation," JCTVC- H0693, San Jose, CA, February 2012.

[6] Y. Syed and D. Holden, "Strategies for Deploying High Resolution and High Frame Rate Cable Content Leveraging Visual Systems Optimizations," Proceeding of NCTA 2012, May 21, 2012.

[7] J.-R. Ohm, G. 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, December, 2012

[8] A. Rodriguez and K. Morse, "HEVC-Driving Disruption in Multiscreen Converged Service Delivery Architecture," Proceedings of 2013 NAB Broadcast Engineering Conference, April 7, 2013, pp 33-38.

[9] ITU-T Rec. H.264 | ISO/IEC 14496-10, (2005), "Information Technology – Coding of audio visual objects – Part 10: Advanced Video Coding."

[10] Gary J. Sullivan, HEVC Tutorial in INCITS meeting at CableLabs, March 2013.