HEVC Demystified

A Primer on the H.265 Video Codec

A white paper from Elemental Technologies

Introduction

The most recent non-proprietary video compression standard, High-Efficiency Video Codec (HEVC), also known as H.265, was placed into final draft for ratification in January 2013 and is expected to become the video standard of choice for the next decade. As with each generation of video compression technology before it, HEVC promises to reduce the overall cost of delivering and storing video assets while maintaining or increasing the quality of experience for the viewer.

HEVC is an open standard, defined by standardization organizations in the telecommunications (ITU-Ts VCEG) and technology industries (ISO/IEC MPEG) to leverage the most efficient video compression techniques using the market's latest processing platforms. Without sacrificing video quality, HEVC can reduce the size of a video file or bit stream by as much as 50% compared to AVC/H.264 or as much as 75% compared to MPEG-2 standards. This results in reduced video storage and transmission costs and also paves the way for higher definition content to be delivered for consumer consumption.

The techniques and algorithms used in HEVC are significantly more complex than those of H.264 and MPEG-2. There are more decisions to make when encoding a given video stream or file and as a result, more calculations need to be made in compressing video assets. This complexity, however, is an excellent fit for video processing solutions that seamlessly evolve from one compression generation to the next as they mitigate the risks that come with any large technological migration.

This paper focuses on the technical and market implications of HEVC's adoption in the content creation and delivery market. It is assumed that in the 1-2 years following ratification of the standard, the vast majority of consumer electronics manufacturers will support hardware-based decoding of HEVC in their devices.

Sponsored by Elemental Technologies

ELEMENTAL

HEVC Demystified
The Broadcast Bridge, August 2014

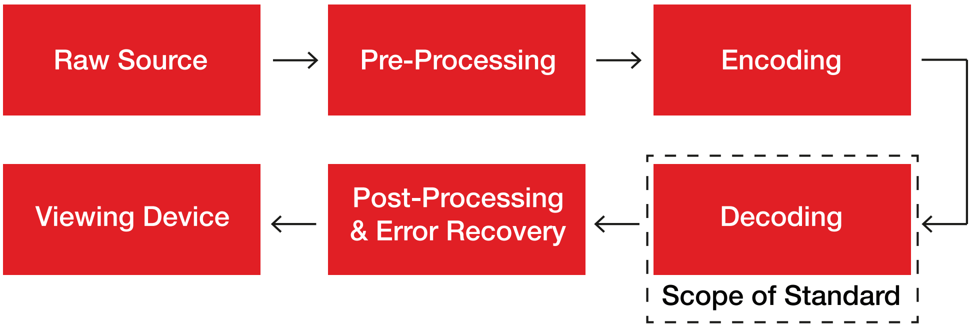

Figure 1. Scope of Video Compression Standardization

Video Compression Basics

Video compression seeks to reduce or remove redundant information from a video stream so that the asset can be stored or sent over a network as efficiently as possible. The algorithms used to eliminate the excess information make up the process of encoding, and the amount of time it takes to accomplish this task is termed “encoding latency,” especially in a linear workflow such as live television delivery. The method used to play back the compressed asset and return it as closely as possible to its original state is known as decoding. Together, the interoperability of the compression and decompression processes form the basis of a codec.

The differences between codecs stem from the techniques each uses to reduce the amount of information in a video bit stream. MPEG-2 achieves different bit rates and quality levels than H.264 and the same is true of HEVC. However, within a given codec standard, the decoder algorithms are firmly defined—in fact, the scope of the standard is generally based around the decoder. In order to achieve interoperability and decode what is known as a “compliant bit stream,” the decoder recipe in figure 1 must be followed.

While encoders must provide the appropriate information for the decoder to follow the specified recipe, encoders within a given standard vary from vendor to vendor, or even from product to product from a single vendor. Encoder variation is caused by how suppliers choose to implement the different sets of tools defined by the standard. Several factors drive which encoding tools are implemented by a vendor, including the needs of its market, the limitations of its particular platform, or decision tradeoffs made by the engineering team designing the encoder.

While a decoder is created using a “recipe”, an encoder may be considered the “secret sauce”. Encoders make the most intelligent approximations and decisions within a field of possibilities that is too vast to fully survey within the constraints of most applications. As a result, not every encoder for a given codec, such as HEVC, will be alike. Encoders do, however, follow these phases:

1. Divide each frame into blocks of pixels so that processing can occur simultaneously at a block level.
2. Identify and leverage spatial redundancies that exist within a frame by encoding some of the original blocks via spatial prediction and other coding techniques.
3. Exploit temporal linkages that exist between blocks in subsequent frames so that only the changes between frames are encoded. This is accomplished via motion estimation and compensation where searches are performed on adjacent frames to create motion vectors that predict qualities of the target block.
4. Identify and take advantage of any remaining spatial redundancies that exist within a frame by encoding only the differences between original and predicted blocks through quantization, transform, and entropy coding.

During the encoding process, different types of video frames, such as I-frames, P-frames and B-frames, may be used by an encoder. When these different frame types are used in combination, video bit rates can be reduced by looking for the temporal (time-based) and spatial redundancy between frames that create extraneous information (see figure 2). In this way, objects, or more precisely, pixels or blocks of pixels, that do not change from frame to frame or are exact replicas of pixels or blocks of pixels around them, can be processed in an intelligent manner.

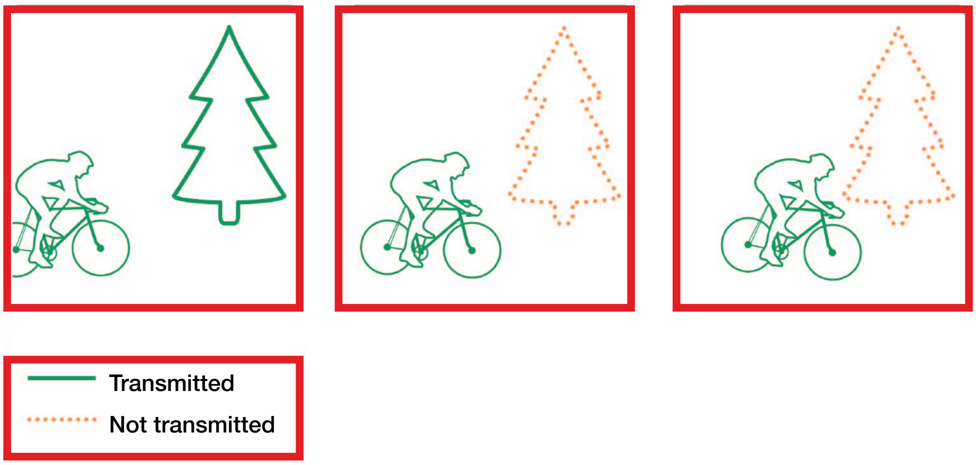

Figure 2. Exploit Commonality and Reduce Data By Identifying The Differences Between frames

With high motion video, such as sports content, the information that stays the same from frame to frame is significantly less important than predicting the motion of objects from frame to frame. With motion compensation algorithms implemented in the encoding process, the codec is able to take into account the fact that most of what makes up a new frame in a video sequence is based on what happened in previous frames. So at a block-by-block level, the encoder can simply code the position of a matching object in the frame and where it is predicted to exist in the next frame via a motion vector (see figure 3). The motion vector takes fewer bits to encode than an entire block and thereby saves bandwidth on the encoded stream.

An **I-frame**, or intra frame, is a self-contained frame that can be independently decoded without reference to preceding or upcoming images.

The first image in a video sequence is always an I-frame and these frames act as starting points if the transmitted bit stream is damaged. I-frames can be used to implement fast-forward, rewind and scene change detection. The drawback of I-frames is that they consume many more bits and do not offer compression savings. On the other hand, I-frames do not generate many artifacts because they represent a complete picture.

A **P-frame**, which stands for predictive inter frame, references earlier I- or P-frames to encode an image. P-frames typically require fewer bits than I-frames, but are susceptible to transmission errors because of their significant dependency on earlier reference frames.

A **B-frame**, derived from bi-predictive inter frame, is a frame that references both an earlier reference frame and a future frame.

A P-frame may only reference preceding I- or P-frames, while a B-frame may reference both preceding and succeeding I- or P-frames.

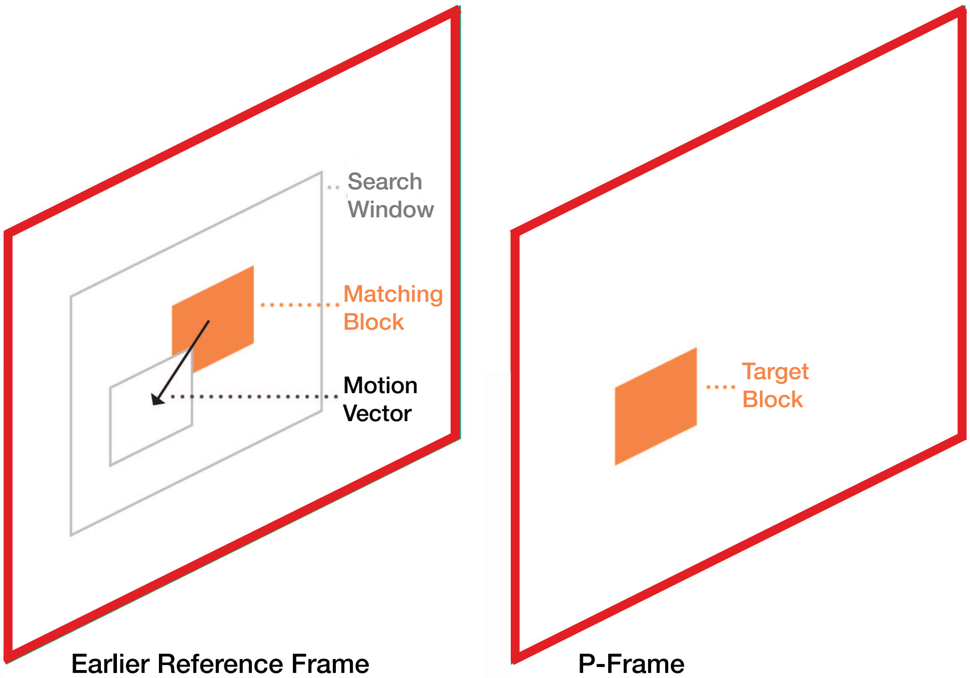

Figure 3. Block-Based Motion Compensation

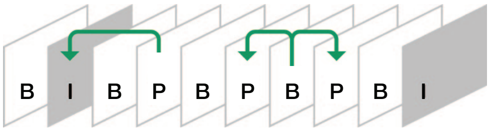

Figure 4. A Typical Sequence with I-, B- and P-Frames

These are the basic techniques and objectives of video compression. There are other algorithms involved that transform information about video into fewer and fewer transmitted bits, but these are not covered in this paper.

Development of HEVC

Like the H.264 standard, HEVC is the output of a joint effort between the ITU-T's Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group (MPEG). The ITU-T facilitates the creation and adoption of telecommunications standards and the ISO/IEC manages standards for the electronics industries. Designed to evolve the video compression industry, HEVC intends to:

- Deliver an average bit rate reduction of 50% for a fixed video quality compared to H.264 (see figure 5)
- Deliver higher quality at same bit rate
- Define a standard syntax to simplify implementation and maximize interoperability
- Remain network friendly—i.e. wrapped in MPEG Transport Streams

The ratified standard of HEVC lays the foundation by defining 8-bit and 10-bit 4:2:0 compression, which applies to the majority of video distribution to connected devices. Further work in the area of 12- and 14-bit encoding at 4:2:2 and 4:4:4 color space is likely to finalize in early 2014.

While H.264 featured seven profiles—a profile is a defined set of coding tools used to create a compliant bit stream—the HEVC spec currently supports three: Main, Main 10, and Main Still Picture. However, the standardization process from January 2013 was only version 1 of the standard and future profile extensions for HEVC will likely include increased bit depth, 4:2:2 and 4:4:4 chroma sampling, Multiview Video Coding (MVC) and Scalable Video Coding (SVC). These extensions are expected in January 2014.

HEVC's Main profile supports a bit depth of 8 bits per color, while Main 10 supports 8 bits or 10 bits per color. Because of the additional bit depth option, Main 10 has the potential to provide better video quality than Main. Finally, Main Still Picture profile allows for a single still picture to be encoded with the same constraints as Main profile.

The HEVC spec also defines 13 levels, which are sets of constraints that indicate the required decoder performance to playback a bit stream of the specified profile. The levels are split into two tiers: Main, which includes levels 1 – 3.1, and High, which includes levels 4 – 6.2 and is designed for highly demanding applications. HEVC levels share a number of similarities with the levels of H.264, with a key difference being the addition of levels 6, 6.1 and 6.2, which define requirements to support 8K resolution video.

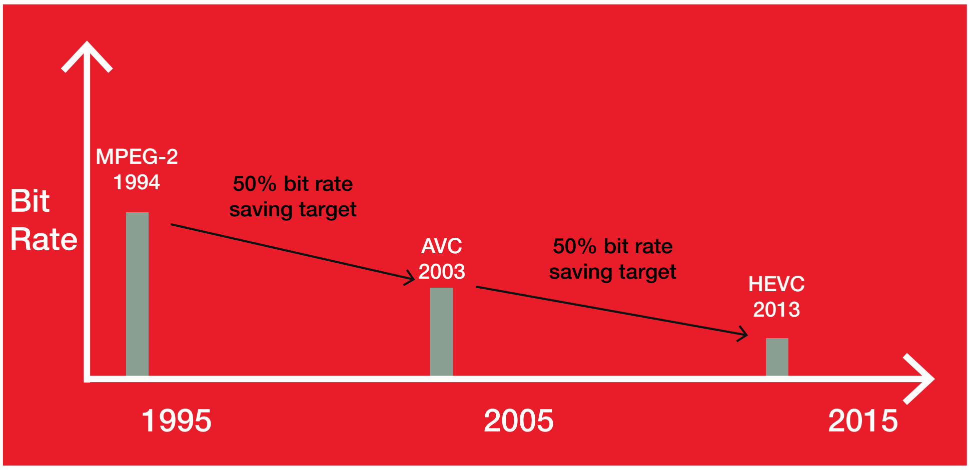

Figure 5. Expected Compression Bit Rates at Time of Standardization

Application Impact

There are several cases where the improved quality to bit rate ratio of HEVC will impact the industry at an application level. As high quality video distribution consumes enormous network capacity, the benefit of these efficiency gains include:

- Deployment of more channels over satellite, cable, and IPTV networks
- Lowered cost of managed and unmanaged video distribution
- Widened reach for bandwidth-constrained mobile and IPTV operators
- Improved QOE of OTT services to match traditional broadcast delivery

In the mobile streaming market, the HEVC bit rate reduction of 30 to 50% to achieve comparable quality to H.264 is realized in the cost savings of delivery across networks. Mobile operators will not need to deliver as much data for a given quality level, making for lower costs and more reliable playback—this, of course, assumes the device's hardware can smoothly decode HEVC.

HEVC also aligns with the push towards high-resolution Ultra HD 4K and 8K video in the mainstream market. With 4K resolution featuring four times the number of pixels as 1080p, the efficiencies provided by HEVC make broadcasting 4K much more feasible.

Media companies with significantly-sized content libraries will also feel the positive impact of bit rate savings. As their storage efforts strive to keep pace with multiscreen consumer demand, these companies face increased strain on their infrastructure. With HEVC halving file sizes, transitioning to the new codec will stretch storage capacity twice as far going forward.

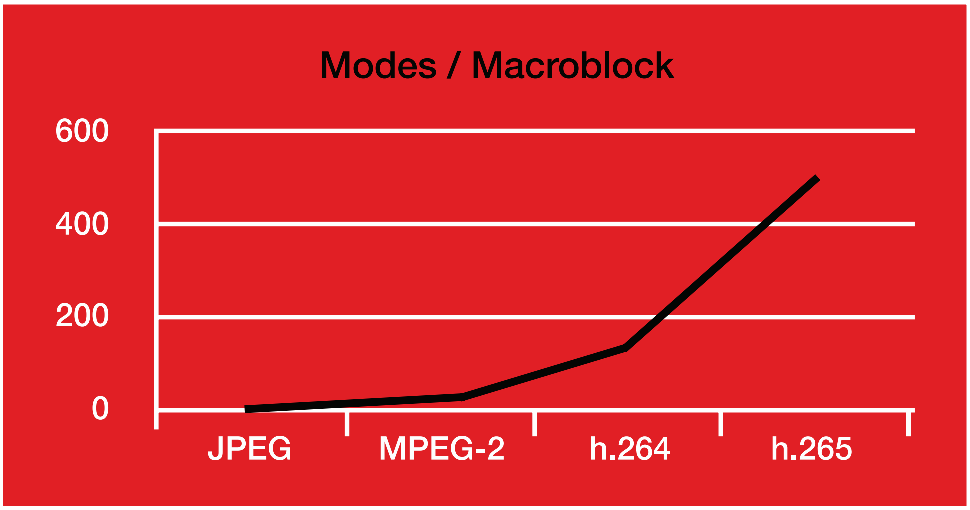

Figure 6. Possible Ways To Encode Each Macroblock

How HEVC is Different

The primary goal of the new HEVC standard is to provide the tools necessary to transmit the smallest amount of information necessary for a given level of video quality. The underlying approach to HEVC is very similar to previously adopted standards such as MPEG-2 and H.264. Simply put: it is much more of the same (see figure 6).

While there are a number of differences between H.264 and HEVC, two stand out: increased modes for intra prediction and refined partitioning for inter prediction.

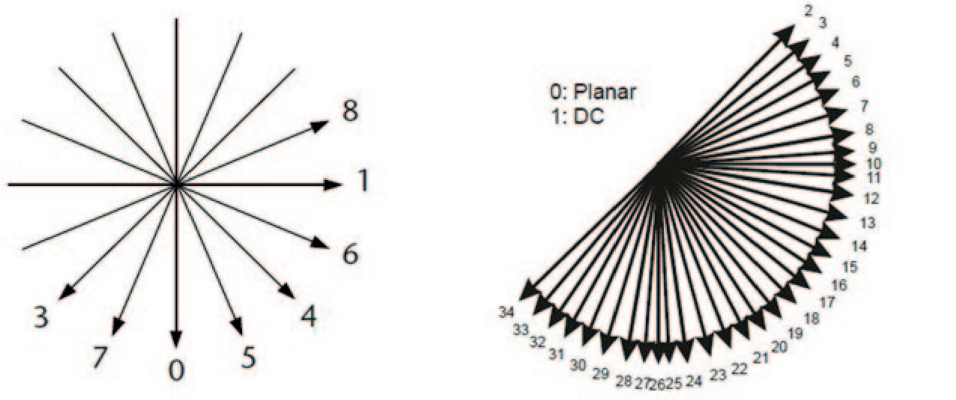

Figure 7. H.264 vs. HEVC Intra Prediction Modes

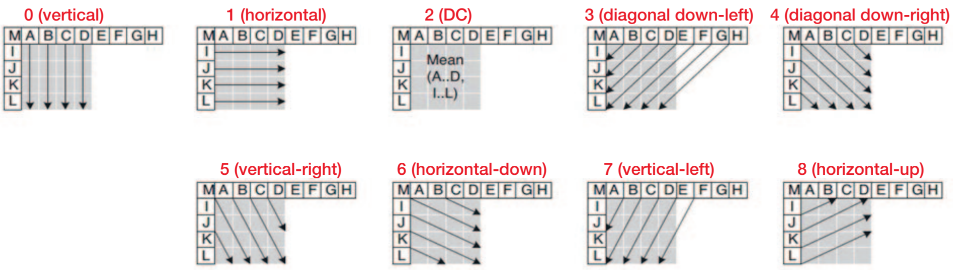

Figure 8. H.264 Intra Prediction Modes in Practice for 4 x 4 Blocks

Intra Prediction and Coding

In the H.264 standard, nine modes of prediction exist in a 4 x 4 block for intra prediction within a given frame and nine modes of prediction exist at the 8 x 8 level. It's even fewer at the 16 x 16 block level, dropping down to only four modes of prediction. Intra prediction attempts to estimate the state of adjacent blocks in a direction that minimizes the error of the estimate. In HEVC, a similar technique exists, but the number of possible modes is 35—in line with the additional complexity of the codec (see figure 7). This creates a dramatically higher number of decision points involved in the analysis, as there are nearly two times the number of spatial intra-prediction sizes in HEVC as compared to H.264 and nearly four times the number of spatial intra-prediction directions (see figure 8).

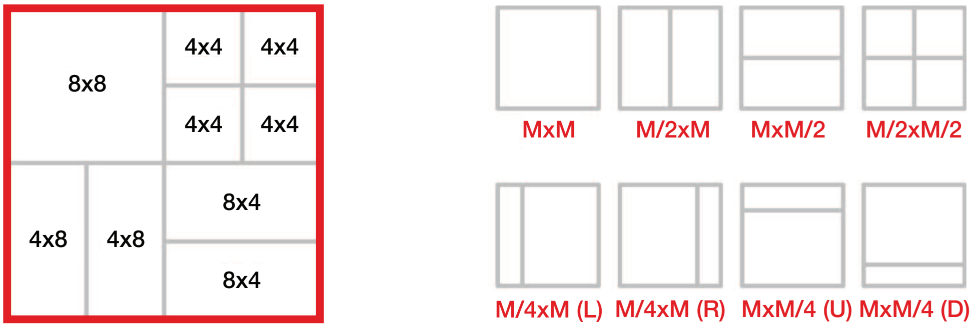

Figure 9. An Example of a 16 x 16 H.264 Macroblock vs. M x M HEVC Partitions

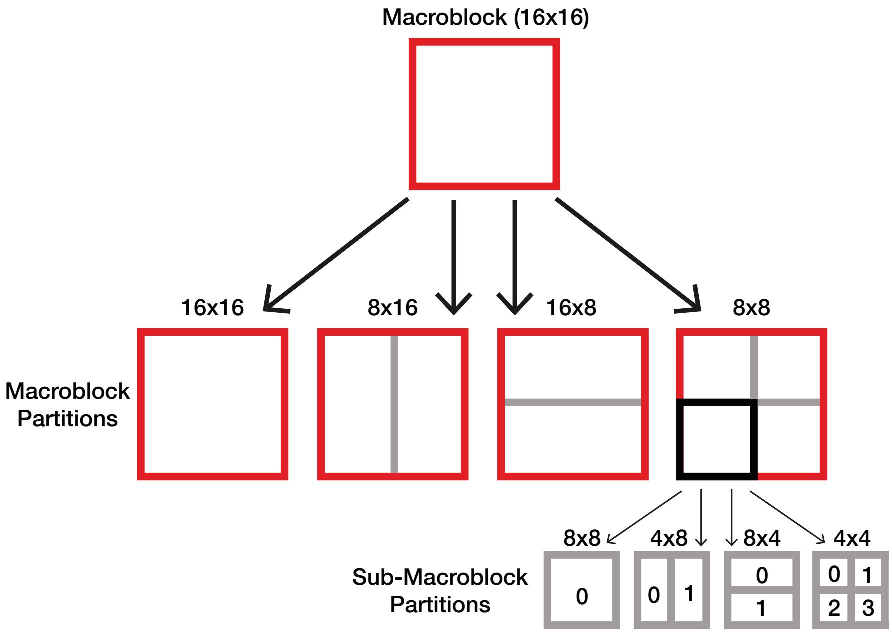

Figure 10. H.264 Macroblock Partitions for Inter Prediction

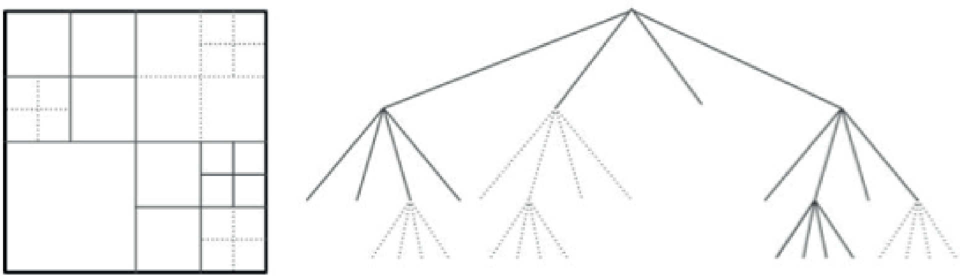

Figure 11. HEVC Quadtree Coding Structure for Inter Prediction

Inter Prediction and Coding

H.264 uses block-based motion compensation with adjustable block size and shape to look for temporal redundancy across frames in a video. Motion compensation is often noted as the most demanding portion of the encoding process. The degree to which it can be implemented intelligently within the decision space has a major impact on the efficiency of the codec. Again, HEVC takes this to a new level (see figures 9 and 10).

HEVC replaces the H.264 macroblock structure with a more efficient, but also complex, set of treeblocks. Each treeblock can be larger (up to 64x64) than the standard 16x16 macroblock, and can be efficiently partitioned using a quadtree (see figure 11). This system affords the encoder a large amount of flexibility to use large partitions when they predict well and small partitions when more detailed predictions are needed. This leads to higher coding efficiency, since large prediction units (up to and including the size of the treeblock) can be cheaply coded when they fit the content. By the same token, when some parts of the treeblock need more detailed predictions, these can also be efficiently described.

HEVC and Parallel Processing

HEVC has been designed keeping in mind improving performance in parallel processing. This includes enhancements for both encoding and decoding. Some of the specific improvements are found in:

- tiles
- the in-loop deblocking filter
- wavefront parallel processing

Tiles allow for a picture to be divided into a grid of rectangular regions that can be independently decoded and encoded simultaneously. They also enable random access to specific regions of a picture in a video stream.

In the case of the in-loop deblocking filter, it has been defined such that it only applies to edges aligned on an 8 x 8 grid in order to reduce the number of interactions between blocks and simplify parallel processing methodologies. Additionally, the processing order has been specified as horizontal filtering on vertical edges followed by vertical filtering of horizontal edges. This allows for multiple parallel threads of deblocking filter calculations to be run simultaneously.

Finally, wavefront parallel processing (WPP) allows each slice to be broken into coding tree units (CTUs) and each CTU unit can be decoded based on information from the preceding CTU. The first row is decoded normally but each additional row requires decisions be made in the previous row.

Elemental and the Fit With HEVC

The computational intensity of HEVC lends itself to the processing performance advantage available with graphics processing units (GPUs). With a flexible software-based architecture, video processing solutions from Elemental offer support for HEVC via a seamless software upgrade. Elemental has deep experience developing video codecs from open specifications to full implementation using general-purpose programmable architectures (GPUs and CPUs). Easing the transition to HEVC within legacy MPEG-2 and H.264 infrastructures, upgradable solutions like those from Elemental can incorporate new compression approaches much more quickly than existing fixed hardware encoding and decoding platforms, such as ASICs and DSPs. Flexible software running on massively parallel hardware makes the Elemental computing platform an ideal fit for HEVC.

While HEVC tools are designed to improve parallel processing capabilities, the sheer number of tools with increased complexity is very large. Elemental has estimated that HEVC encoding will require up to ten times more processing power than H.264 encoding. For example, with 500 different ways to encode each macro block, the processing power requirements are significantly higher for HEVC when compared with H.264 encoding. A doubling of spatial intra-prediction sizes, doubling of the number of transform sizes, almost four times the number of spatial intra-prediction directions, and the increase in the inter prediction search space will strain many existing hardware platforms and highlights the value of powerful GPU-based video processing in media companies' endeavors to support the new codec.

The introduction of ultra-high resolution 4K and 8K content to the consumer market coupled with the focus on HEVC to support these resolutions will cause additional issues for most existing encoding solutions. GPUs hold an advantage again in this case, as these powerful processors are optimized to handle increased resolutions.

Conclusion

It has not been very long since the migration from MPEG-2 to H.264, and the H.264 to HEVC transition will require a similar re-architecting of hardware and software. From a software point of view, Elemental is one of very few companies in the world with experience implementing multiple video codecs from specification to full implementation. These implementations have been performed using hybrid GPU/CPU platforms with standard software tools to optimize for both quality and performance. This resulting computing platform creates an environment that not only meets the performance requirements of HEVC but is also programmable and simple to upgrade, as required by evolving specifications and emerging technologies.

HEVC represents a significant move forward in video compression technology. It harnesses an expanded set of tools and algorithms as well as improved parallel processing capabilities to enable better compression efficiencies. These compression efficiencies allow media companies to create higher quality video streams with larger resolutions at the same bit rates of previous generation codecs, reducing the overall cost of video asset delivery while focusing on the quality of experience for the viewer. Elemental expects that HEVC will find widespread adoption in streaming, broadcast, satellite, cable, IPTV, surveillance, corporate video and gaming applications before the end of the decade.

COMPONENT	MPEG-2	H.264	HEVC/H.265
General	Motion compensated predictive, residual, transformed, entropy coded	Same basics as MPEG-2	Same basics as MPEG-2
Intra prediction	DC Only	Multi-direction, multi-pattern, 9 intra modes for 4x4, 9 for 8x8, 4 for 16x16	35 modes for intra prediction, 32x32, 16x16, 8x8 and 4x4 prediction size
Coded Image Types	I, B, P	I, B, P, SI, SP	I, P, B
Transform	8x8 DCT	8x8 and 4x4 DCT-like Integer Transform	32x32, 16x16, 8x8 and 4x4 DCT-like Integer Transform
Motion Estimation Blocks	16x16	16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4	64x64 and hierarchical quad-tree partitioning down to 32x32, 16x16, 8x8 Each size can be partitioned once more in up to 8 ways
Entropy Coding	Multiple VLC tables	Context adaptive binary arithmetic coding (CABAC) and context adaptive VLC tables (CAVLC)	Context adaptive binary arithmetic coding (CABAC)
Frame Distance for Prediction	1 past and 1 future reference frame	Up to 16 past and/or future reference frames, including long-term references	Up to 15 past and/or future reference frames, including long-term references
Fractional Motion Estimation	½ pixel bilinear interpolation	½ pixel 6-tap filter, ¼ pixel linear interpolation	¼ pixel 8-tap filter
In-Loop Filter	None	Adaptive deblocking filter	Adaptive deblocking filter and sample adaptive offset filter

Figure 12 Table. MPEG-2 vs. H.264 vs. HEVC

References

Overview of the H.264/AVC Video Coding Standard, Thomas Wiegand, Fellow IEEE, Gary J. Sullivan, Senior Member, IEEE, Gisle Bjøntegaard, and Ajay Luthra, Senior Member, IEEE.

Overview of the High Efficiency Video Coding (HEVC) Standard, Thomas Wiegand, Fellow IEEE, Gary J. Sullivan, Fellow IEEE, Woo-Jin Han, Member IEEE, and Jens-Rainer Ohm, Member IEEE.

http://people.irisa.fr/Olivier.Le_Meur/teaching/HEVC_CAV_ESIR3_2011_2012.pdf

http://www.h265.net/