H.264 vs H.265

Malcolm Weir

George Nelson

Gary Thom

Delta Information Systems, Horsham, PA, USA
Ampex Data Systems Corporation, Hayward, CA, USA

Abstract: This paper discusses the application of two of the latest video encoding approaches for test platforms, particularly in bandwidth-constrained and over imperfect (lossy) transmission links. High Efficiency Video Coding (HEVC or H.265) is compared with MPEG-4 Advanced Video Coding (AVC or H.264), and the behavior of both in low bandwidth and lossy transmission channels is explored, concentrating on the ability to transmit usable information over low bandwidth links using each approach. Comparison of the same video simultaneously compressed heavily for constrained links and lightly for on-board storage is used to illustrate the effects of heavy compression on video usability, as well as side-by-side comparison of the output of both video encoding algorithms as implemented in a rugged airborne package.

Keywords: Video, Encoding, CODEC

1. Introduction

The latest video encoding standard, High Efficiency Video Coding (HEVC), from the body[1] that developed the H.264/MPEG-4 AVC standard was published in June 2013. As with its predecessor, the standard has two names reflecting the two standards bodies: ITU-T's H.265 and ISO/IEC MPEG-H Part 2. In general, the marketplace seems to have adopted H.265/HEVC as a preferred nomenclature.

Historically, it has taken roughly five years or so for a new video coding standard to gain traction. By many measures, H.265/HEVC is running ahead of that, with substantial uptake in the commercial world for distribution of video over networks. Amazon Prime, Netflix and the BBC all use H.265/HEVC encoding for some of their content.

While part of the reason for the adoption of the new encoding format is the added support for Ultra High Definition (UHD) TV formats such as "4K", another motivator is the reduced bandwidth required by lower-resolution formats. Clearly, if there exists a mechanism suited to deliver the "4K" format 2160p60, which is eight times the number of pixels per second required by 1080p30, then that same mechanism could be applied to less demanding formats (e.g. 1080p30) and achieve some level of bandwidth savings.

It is this second rationale that has led to a category of applications for H.265/HEVC that is totally distinct from mass content distribution (as implemented by the likes of Netflix for UHD-TV): video transmission over constrained network links, as with Apple's FaceTime application connecting iPhone 6 (and later) devices.

Therefore, with H.265/HEVC in use within a ubiquitous mass-market smartphone, it is clear that here is a technology that may be ready for use in test and evaluation applications.

In the related market of airborne ISR technology, the US Department of Defense's Motion Imagery Standards Board (MISB), and the related STANAG (STANAG 4609) has provisionally endorsed the use of H.265/HEVC. As the entity responsible for ensuring interoperability of video systems, this provides a strong indication that the technology is stable and mature.

2. Commercial Aspects of H.265

While the technical merits of any video encoding method will obviously inform the decision as to whether or not to use the technology, other factors will also influence the process.

2.1. Patents and Licensing

Like its predecessor, H.264, several keys parts of the H/265/HEVC technology are the subjected of patents. However, again like H.264, there is a significant threshold before any royalties are due: the first 100,000 units sold each year attract no royalties[2]. So for all intents and purposes, there seems to be little risk of a commercial impact from patent costs.

2.2. Alternative Technologies

While considering the merits of H.265/HEVC, it is worth also noting that there is a potential alternative sponsored by

---

[1] *The Joint Collaborative Team on Video Coding (JCT-VC), comprising the ISO/IEC Motion Picture Expert Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG).*

[2] *This refers to the largest patent pool administered by MPEG-LA; other patents also apply, but similar considerations seem to cover these.*

Google: the VP9 coding scheme. By contrast to the the ISO/ITU-T methods, the intent is that VP9 is will be unencumbered by patent claims; whether this intention can survive a legal challenge (should one be made) remains to be seen. However, no assertion of infringement has been made, and indeed political (anti-trust) considerations have resulted in what appears to be a truce between the various rights holders.

From a technology perspective, VP9 is designed for a different use case; while H.265/HEVC is largely driven by the requirements of "4K" video distribution (and can be used with lower resolutions), VP9 is the inverse, being primarily motivated to improve the compressibility of "1080p" video, but also able to be used with larger frame sizes. Given the sponsorship by Google, it is obvious what motivates the design: Google's YouTube.com is source of a huge volume of traffic, so if they can reduce that volume in any meaningful way, then a huge savings in total traffic is achievable.

However, the authors believe that, whatever the respective merits and drawbacks of H.265/HEVC versus VP9, the former will dominate based primarily on the breadth of support. A survey of the various implementations of each method show that there are simply more solutions that encode and/or decode H.265/HEVC than there are for VP9. Hence from a commercial standpoint, it seems reasonable to assume that, absent a specific compelling technical reason, H.265/HEVC is the preferred "successor" to H.264/AVC.

3. Architecture of H.265 compared with H.264

In most respects, H.265/HEVC is a direct evolution of H.264/AVC. While many of the individual techniques are refined and improved in H.265/HEVC, there is an obvious correspondence between the two coding methods.

3.1. Specifications

The general design specifications of H.265/HEVC and H.264/AVC are outlined in Table 1.

Feature	H.264/AVC	H.265/HEVC
Compression Model	Hybrid Spatial/Temporal Prediction	Enhanced Hybrid Spatial/Temporal Prediction
Frame Size (Max)	4K / 2160p (4,096 x 2,304)	8K / 4320p (8,192 x 4,320)
"Sweet Spot" Video Format	1080p30	2160p60
Frame Rate	59.94 fps	300 fps
Interlaced Modes	Yes	Reduced
Target Bandwidth	50% of MPEG-2/H.262	50% of H.264/AVC

Table 1 Specifications Compared

(Note: the ability for H.265/HEVC to code interlaced video streams – such as 1080i – is retained, but the implementation is by selectively encoding either de-interlaced frames or the individual fields, and switching between the two approaches on the fly).

3.2. Transports and Containers

Both H.264/AVC and H.265/HEVC use the same structures to stream and store video; in this regard they are interchangeable, which means that audio and other information – most notably, KLV metadata – is unaffected by a change from H.264/AVC to H.265/HEVC. In fact, the design of the MPEG Transport Stream permits a single stream to carry both, simultaneously, as two different programs, as shown in Figure 1. (Note that this makes sense in relatively few use cases, and is included simply to indicate the possibility. Whenever it is desirable to send two streams of encoded video / audio / metadata, it is usually preferable to generate two distinct and independent Transport Streams).

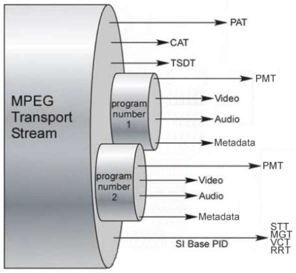

Figure 1 MPEG Transport Stream

Naturally, H.265/HEVC encoded video can simply be saved as an MPEG Transport Stream in a file, but it also can be used within the "Matroska" (.MKV) and ".MP4" containers. It is not, however, supported by the ".AVI" or "QuickTime" formats; this is mostly an indication that those containers types are largely deprecated (rather than any technical issue related to the newer encoding scheme) .

This commonality with H.264/AVC means that H.265/HEVC can be managed in the same way, which in turn means that organizations, particularly including the MISB, have granted approval to use H.265/HEVC as an alternative to the older format[1].

3.3. Encoding Techniques

Both H.265/HEVC and H.264/AVC use the same general approach to encoding a video sequence: the initial frame in a sequence is encoded using only intra-picture methods, and then subsequent frames use both intra-picture and inter-picture compression.

The most obvious difference between the two schemes is that H.265/HEVC replaces 16x16 pixel macroblocks with a new construct, called a Coding Tree Unit (CTU). CTU's can be as large as 64x64 pixels, but critically within a CTU the image is further divided into one or more Coding Blocks (CBs). The size of a CB depends on the level of detail in an image. This, simply put, it allows the algorithm to use large blocks where there's not much detail (e.g. an expanse of wall) and small blocks where there is more detail (e.g. someone's face).

Within each CB, additional techniques (compared to with H.264/AVC's macroblocks) are available to encode the image. The effectiveness of this technique is illustrated by the fact that forcing an H.265/HEVC encoder to use only 16x16 pixel CTU's increases the output bit rate by more than 25%.

The next improvement lies with intra-frame prediction; H.264/AVC uses eight angular prediction modes, while

H.265/HEVC increases the number of angular modes to thirty-three.

Motion Vector (MV) prediction, used for inter-picture compression, is likewise enhanced; MV's use larger 16 bit values, which translates to an ability to represent four times the offset (+/- 8192 vs. +/- 2048 horizontally and +/- 512 vertically).

Clearly, all these additional variations add dramatically to the processing power required to encode a video sequence; depending on the sophistication of the encoder, anywhere from 300% to 1000% increase in CPU requirements together with significant memory accesses (in the order of gigabytes per second) have been measured. To help offset the real-world consequences of these increases, H.265/HEVC includes provision for using massively parallel algorithms (and hardware), including the ability to "tile" a frame and measures to allow distinct threads to synchronize themselves along a "wavefront". The deblocking filter has also been adapted (actually, made less flexible) to support greater parallelization.

Finally, a new filter, the Sample Adaptive Offset (SAO) filter, is applied to help reduce the appearance of common artefacts, such as banding and ringing.

4. Quality vs Bandwidth

Image quality can be assessed in terms of an objective, mathematical calculation, typically the Peak Signal-to-Noise Ratio (PSNR). However, contemporary video encoding schemes are designed to exploit the perceptual limits of the human visual system and brain, so objective measures can be misleading; one well known example of this phenomena is with MP3 audio compression, which is deliberately "tuned" to the human ear, so it "sounds better" than the PSNR measure would seem to indicate.

The better approach is to use a subjective analysis employing multiple real people in as controlled an environment as is practical, and then "average" their subjective ratings to create a useful comparison tool.

This sort of experiment is, of course, costly and time consuming to manage. Fortunately, the British Broadcasting Corporation (BBC) Research and Development Department created and executed a test plan, with the assistance of the JCT-VC, and provided the results in *Tan et al.*[2].

4.1. Test Setup

The experiments performed by the BBC used twenty different video sequences in a number of different resolutions: 2160p, 1080p, 720p and 480p. For each sequence, the subjects saw both the original uncompressed video, and then the compressed video at a number of different bitrates for both H.265/HEVC and H.264/AVC.

For each sequence and bitrate, both the PSNR and the "Mean Opinion Score" (MOS) from the subjective results were calculated, and the results plotted on a chart such as the one shown in Figure 2.

From this chart, a well-accepted model for evaluating coding efficiency, the Bjøntegaard model[3] was used to calculate a "Bjøntegaard Delta" metric for the bit rate, expressed as a percentage of the reference (i.e. H.264/AVC) rate.

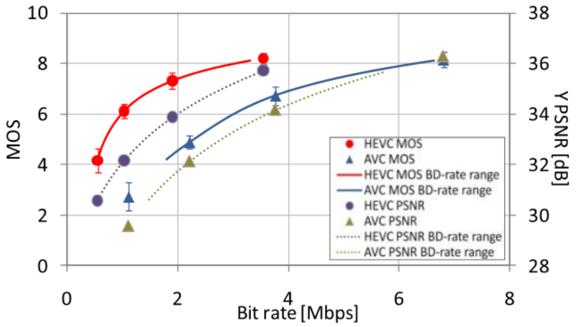

Figure 2 Video Comparison Chart

4.2. Results

The condensed version of the BBC's extensive analysis is summarized in Table 2 below.

Sequence	Bitrate Reduction for Equivalent Quality	
	Objective BD-Rate (PSNR)	Subjective BD-Rate (MOS)
2160p	47%	63%
1080p	50%	61%
720p	42%	56%
480p	39%	57%
Average	**44%**	**59%**

Table 2 Bitrates for Equivalent Quality

The first, and most obvious, conclusion is that the H.265/HEVC coding scheme does indeed deliver on it's design goal: for a given image quality, approximately half the bit rate is required compared to H.264/AVC.

The second conclusion is a little more subtle: based on the difference between the subjective and objective figures, H.265/HEVC can be considered to be better optimized for the human brain than is H.264/AVC. Or, to put it another way, H.265/HEVC uses bits more wisely, directing greater detail to the visually more important areas of a video.

5. Quality vs Channel Loss

In the Test & Evaluation (T&E) community, it is sometimes a challenge to ensure reliable delivery of streaming data; many test systems use telemetry links that are prone to dropouts as the test vehicle moves through space. H.265/HEVC provides a couple of capabilities that help deliver quality better than H.264/AVC under such conditions.

5.1. Error Correction

The simplest approach to managing lossy communication channels is to simply use some of the "saved" bandwidth to add sufficient error correction structures to overcome the anticipated losses. One scheme for this if the communication channel is unidirectional is the forward error correction in IRIG 106 Chapter 7, "Packet Telemetry Downlink"; for bidirectional links, the issue typically gets resolved by retransmission of lost or corrupted packets.

5.2. Subjective Consequences

Perhaps more than with any other situation, channel loss is measured by the impact to the consumer, i.e. the subjective quality, as opposed to objective measures. For instance, if one can compensate for a lossy channel with error correction, then, from a subjective perspective, the losses are invisible and simply appears to have a lower bandwidth.

With subjective considerations, H.265/HEVC offers definite advantages over H.264/HEVC. First, as the BBC experiment shows, the subjective analysis (MOS) of the video samples shows that comparable video quality is obtainable with more than the 50% bandwidth reduction targeted by the design. Therefore, even if only half the bandwidth (of H.264/AVC) is allocated for a video stream, there is additional "headroom" for managing the lossy channel. Alternatively, because the subjective quality of the video at 50% of the H.264/AVC bitrate is "better" than that of the H.264/AVC video, distortion introduced by lost data will have a lower subjective impact than the same loss at double the bitrate using H.264/AVC.

Secondly, the design of the H.265/HEVC Coding Tree architecture means that, considered across a single frame, in the encoded version, more bits will be used for detail (e.g. an instrument) than for background (e.g. the sky). So losing one packet's worth of data (up to 182 bytes) is, statistically, more likely to result in a loss of data in a smaller, rather than larger, piece of the image. Subjectively, then, data loss tends to be less visually intrusive and thus of greater utility than would have been the case with older encoding methods.

Thirdly, because the H.265/HEVC coding uses half the bandwidth compared to H.264/AVC, it is practical to use higher resolutions and/or better quality settings than would have been the case with the earlier method. If the quality setting is (simplistically) likened to the "fuzz" reducing the clarity of a given picture, then reducing that "fuzz" or increasing the detail subject to "fuzz" both result in increasing the ability of the human brain to resolve detail (for example, distinguishing between the numerals "0" and "8").

6. Additional Test & Evaluation Considerations

Commercial broadcasting, which is undeniably the primary motivator for most video coding technologies, differs significantly from the CONOPS of Test and Evaluation applications.

6.1. Video "Workflow"

Broadcasting tends to be oriented around a workflow consisting of three distinct phases: Capture, then Edit, and finally Disseminate. Sometimes, certainly, the "Edit" phase is performed live (e.g. with sports broadcasts), but overall the "consumption" of the video is separated from the "production" of it. This separation encourages a model whereby the encoding for the ultimate consumer is distinct from the acquisition phase.

Test & Evaluation applications, in contrast, tend to closely link the producer and consumer, which implies that the coding of video for the consumer will happen at the point of acquisition. This would be unremarkable, except that T&E *also* contends with two competing use cases: high quality (for analysis of anomalies) and low bandwidth (for telemetry).

The "broadcasting" approach to that situation would be to acquire data using a very quality / high bitrate configuration, and then transcode the video *after* the Edit stage to a lower quality and bitrate. Obviously, for a T&E application that is suboptimal for multiple reasons, including the added latency and the power and space required for the transcoding equipment. Further, if (as is common) there is a desire to simultaneously record data on-board while telemetering data to the ground, another transcoding effort is probably appropriate, so the acquisition, recording and telemetry video streams would all use different quality / bitrate settings; for example, the MISB "Profiles" define different "levels" for Acquisition, Processing/Archiving and Distribution[4].

The simplest and most effective approach for T&E applications to resolve this situation is to produce multiple compressed outputs from a single uncompressed input. This can be achieved with a "daisy chain" approach, but far more efficient is a scheme that uses a single video acquisition stage in front of two or more compression engines.

While this design is not unique to H.265/HEVC, the significantly reduced additional cost in terms of power consumption, etc. increases the impetuous for combined encoders, while the practicality of telemetering an full H.265/HEVC stream increases the likelihood of a compromise in quality/bitrate that satisfies no-one.

A further refinement derives from the fact that, as H.265/HEVC is substantially more computationally and memory intensive to encode than H.264/AVC, a sophisticated encoder could produce *both* H.265/HEVC *and* H.264/AVC outputs from the same input. This might be useful if it is desirable to retain the H.264/AVC format to suit an established processing workflow while adopting H.265/HEVC for a telemetry link. Typically, the H.264/AVC would be lightly compressed and stored on the test article, while the H.265/HEVC stream would be encoded with the link characteristics (bandwidth, signal loss, etc) in mind.

As noted in section 3.2 above, the outputs – the compressed video streams – can be combined in a single MPEG Transport Stream, but possibly more practical would be produced and managed as two distinct network streams, so that normal networking techniques, such as multicasting, routing, encryption etc. can be selectively applied to the various outputs.

6.2. High Frame Rates

As indicated in section 3.1, H.265/HEVC can support frame rates up to 300fps. While that is not sufficient for true high-speed video (typically 1000fps and up), it does open up a significant new opportunity for conventional video systems to provide new tools for T&E applications.

7. Conclusion

When digital video first became a mainstream technology[3], 4-6Mbps were required to transmit a high-quality, standard definition, MPEG2 encoded video stream. With H.265/HEVC, that bandwidth is more than enough for at least one very good quality high definition signal, or more signals with judicious compression settings.

H.265/HEV has absolutely delivered on its promise to use half the bandwidth of an equivalent H.264/AVC stream, and has introduced techniques that have improved its subjective quality even further.

Those factors have pushed the adoption of H.265/HEVC faster than has been predicted, and has lead to the conclusion that the time is ripe for the adoption of the new standard.

8. Acknowledgements

The authors acknowledges with thanks the openness of the flight test community, in both industry and government across the civil and military marketplace, and their willingness to share not only the challenges of their work but also to entertain wide-ranging potential solutions to those challenges.

The authors would also like to thank their colleagues and predecessors at Delta and Ampex for the innovations that have helped make video handling in harsh environments a solvable problem.

9. References

[1] See http://www.gwg.nga.mil/misb/faq.html#section2.2, retrieved April 2017.

[2] Tan *et al.* "Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance", IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), January 2016. See http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7254155. Retrieved April 2017.

[3] G. Bjøntegaard, Calculation of Average PSNR Differences Between RD-Curves , document VCEG-M33, ITU-T SG 16/Q6, 13th VCEG Meeting, Austin, TX, USA, Apr. 2001.

[4] Motion Imagery Standards Profile, see http://www.gwg.nga.mil/misb/misp_pubs.html, retrieved April 2017.

10. Glossary

AVC:	Advanced Video Coding
BBC:	British Broadcasting Corporation
CTU:	Coding Tree Unit
DOD:	Department of Defense (US)
HEVC:	High Efficiency Video Coding
IEC:	International Electrotechnical Commission
ISO:	International Standards Organization
ITU-T:	International Telecommunication Union – Telecommunication Standarization Sector
JCT-VC:	Joint Collaborative Team on Video Coding
MISB:	Motion Imagery Standards Board
MOS:	Mean Opinion Score
MPEG:	Motion Picture Experts Group
PSNR:	Peak Signal-to-Noise Ratio
SMPTE:	Society for Motion Picture and Television Engineers
T&E:	Test and Evaluation
UHD:	Ultra High Definition

---

[3] *Circa 1996, when Digital Versatile Disks (DVD) became available. Bit rates for DVD movies are typically 4-5Mbits/sec, with an absolute maximum of 9.8Mb/s.*