

Next-Generation Video Compression Techniques

By Thierry Fautier

Introdução: A compressão de vídeo ainda é um assunto muito relevante, especialmente se a indústria deseja abordar novas aplicações, dominar definitivamente as redes móveis e garantir presença em países emergentes. Este paper descreve como novas topologias baseadas em rede podem trazer grandes melhorias além da compressão de codec pura, não apenas para novos codecs, mas também para os existentes como HEVC, AV-1 entre outros. Para isso The Motion Picture Experts Group, (MPEG para os íntimos) estão trabalhando nos requisitos de um novo codec, de geração futura que deverá fornecer reduções de taxa de bits de 50% em comparação com o HEVC para a mesma qualidade perceptiva. Uma boa leitura a todos!

Por: Tom Jones Moreira

Abstract

The Motion Picture Experts Group is working on requirements for a new video codec, expected to be completed by 2020. The future-generation codec is anticipated to provide a 50% bitrate reduction compared with the High Efficiency Video Coding (HEVC) main profile for the same perceptual quality. Although this is sufficient for certain use cases, it may justify a future video coding standard for other use cases that require bitrate reductions higher than 50%. This paper focuses on target applications for the new video codec that includes 8K, virtual reality (VR) in 3DoF and 6DoF, the state of IP networks, and better compression through Network Distributed Video Coding (NDVC). The author elaborates on how new techniques leveraging content-aware encoding, elastic encoding, machine-learning techniques, and pre- and post-processing pairing can lead to factor four compression efficiency. This paper also compares different approaches including HEVC, HEVC enhanced with NDVC techniques, and the Joint Video Exploration Team codec compared with NDVC techniques.

Keywords

AV-1, bitrate, codec, High Efficiency Video Coding (HEVC), Joint Video Exploration Team (JVET), Motion Picture Experts Group (MPEG), video compression, VP-9

Introduction

State of the Union on Codecs

HEVC. The High Efficiency Video Coding (HEVC) codec has always been a topic of intense research and competition. Motion Picture Experts Group (MPEG) has led the race with its most recent incarnation HEVC, deployed already in ultrahigh-definition (UHD) (i.e., Blu-ray, Advanced Television Systems Committee 3.0 broadcast, and over-the-top (OTT)) and HD broadcast (i.e., Digital Video Broadcast standardization (DVB-T-2)) applications. In the future, it is expected to be used in other bandwidth-sensitive applications such as OTT, especially over

mobile networks. Despite its widespread use, the licensing model is an issue. This has led to the creation of the Open Media Alliance, potentially creating a codec war.¹

AV-1. AV-1, the codec developed by the Open Media Alliance, is expected to provide 40% bitrate savings over VP-9, which was benchmarked by Netflix at 20% higher bitrate than HEVC at the same video quality level. This would make AV-1 a middle of the road solution for OTT in 2017, with commercial deployments in 2018.

JVET. ISO/MPEG and International Telecommunication Union-T/Video Coding Expert Group have announced a new joint codec initiative called *Joint Video Exploration Team (JVET)*. The call for proposals is expected to occur in October 2017. Based on the MPEG requirements for a future video coding standard, the goal is to deliver a 50% bitrate reduction over HEVC by the end of 2020.

Figure 1 shows the expected state of codecs by 2020. As there is no formal comparison made by one single entity on the same test pattern, with no agreement about quality acceptance criteria, those numbers have a very high variance compared with formal tests done inside MPEG.

This paper focuses on target applications for the new video codec that includes 8K, VR in 3DoF and 6DoF, the state of IP networks, and better compression through Network Distributed Video Coding (NDVC).

Bitrate Requirements

This section defines all of the bitrate requirements as of today and what can be expected by 2020.

Existing Applications

Table 1 provides the nominal bitrates associated with existing video applications used for streaming and broadcast. The numbers provided are an aggregation across different commercial deployments. The numbers should be considered representative, accurate, and realistic.

The following is a short description of each format:

- standard definition (SD) is most commonly used on constrained fixed networks (Digital Subscriber Line) or mobile networks.
- HD30 is the sweet spot of OTT on non-TV displays.
- 1080p60 is the best tradeoff, in terms of quality and bitrate, on TV displays.

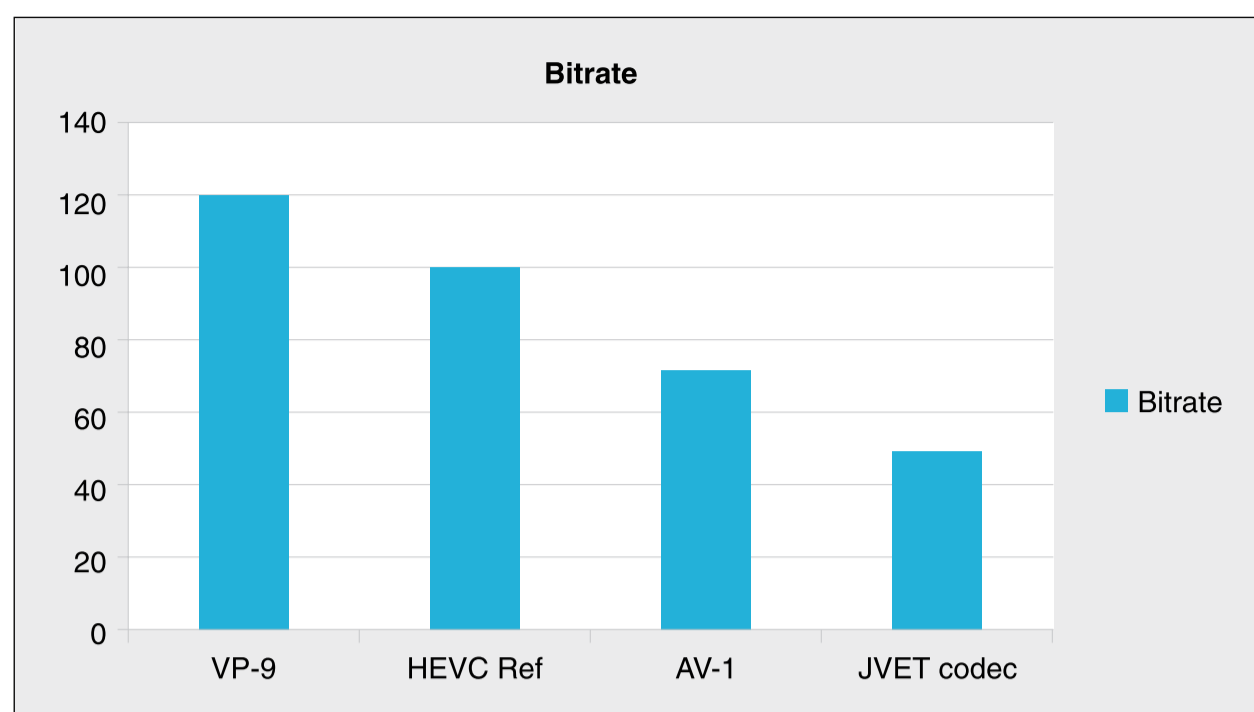
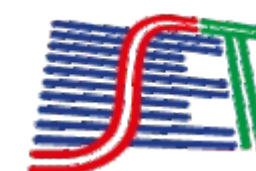


FIGURE 1. Codec comparisons.

- virtual reality (VR) legacy is the current maximum streamed resolution to various head-mounted displays (HMDs). Note: In all deployed systems, the full resolution is sent to the HMD, although only a subregion is being watched (approximately one-tenth), meaning it is a suboptimal system.
- UHD-1 is the broadcast or OTT best quality of experience. Note: For adaptive streaming, subresolutions have to be applied.

Next-Generation 2020 Applications

Table 2 shows the expected bitrates associated with new video applications used for streaming and broadcast. The numbers provided are predictions of the future performance of technologies as well as the type of services that are expected to be launched in the coming years. This is assuming that there is a yearly compression gain of 10% between 2016 and 2020 for HEVC, across all resolutions and bitrates. Note: The numbers provided are projections; therefore, they may not be accurate and could vary by 2020.

The following is a short description of each format:

- VR tiling is a scheme where only the watched section is shown on the HMD, resulting in sub-UHD resolution.
- VR tiling HD means that capture is done in UHD resolution. Only a set of tiles of HD combined resolution is sent to the HMD. Thus, it achieves the same quality as VR Legacy, but divides the bitrate by three.
- VR tiling UHD is when capture is done at approximately 8x UHD-1 resolution, and only a set of tiles

of UHD combined resolution is sent to the HMD. This achieves a much higher quality versus VR legacy, as ten times more pixels are presented on the HMD.

- VR tiling UHD-1 120 is when capture is done at approximately 8x UHD-1 120 resolution, and only a UHD-1 120 set of tiles is sent to the HMD. It achieves a much higher quality versus VR tiling UHD, as the refresh rate is increased to 120 frames/s, which is very important to prevent nausea and decrease the tiling refresh perception.
- UHD-1 120 is the best possible quality that can be achieved for UHD in an economical way by 2020, especially for sports and fast action.
- UHD-2, also known as 8K, is the format NHK promotes. It will be demonstrated as a service at the 2020 Olympic Games in Tokyo.
- VR 6DoF uses volumetric capture, and one frame is 3 GB. If we assume the same compression used for UHD, this would lead to 1.5 Gbits/s, which is a better fit for download/storage media than streaming.

Conclusion

With the increase in resolution provided by VR, thanks to the techniques such as tiling where only the field of view is transmitted, the bandwidth impact can be included. In applications such as UHD, where the resolution is expected to increase from 4K to 8K, the bandwidth loses its control as it reaches bitrates as high as 65 Mbits/s using HEVC. VR 6DoF is clearly out of transmission boundaries and will be considered to be more of a storage media application. The next section will review, in more detail, how those requirements correlate with the application and the network types.

Application Requirements

MPEG maintains a living document that lists the applications and provides the impact on the codec itself.² This gives a broad overview about the function of the video and the types of supported applications. At the first MPEG roadmap session that took place in Chengdu, China,³ multiple institutions (i.e., regulators, research institutes, and operators) presented their views for 2020, and it was clear that UHD-2 (8K), VR, and OTT were the three major applications highlighted.

TABLE 1. Bitrates for existing applications.

Format	Application	Transmitted resolution	Bit depth	Codec	Bitrate (Mbit/s)
SD	Streaming	720 × 480 × 30	8 bit	AVC	1.5
HDp30	Streaming	1280 × 720 × 30	8 bit	AVC	3
1080p60	Streaming/broadcast	1920 × 1080 × 60	8 bit	HEVC	4
VR Legacy	Streaming	2560 × 1440 × 30	8 bit	HEVC	15
UHD-1	Streaming/broadcast	3840 × 2160 × 60	10 bit	HEVC	20+



Video Share in Internet Traffic

This section reviews the origin of internet video traffic.

UHD

UHD-2 that is currently encoded at around 100 Mbits/s, in HEVC, might be done on a wideband transponder with DVB-S2X; however, it could never fit economically on a terrestrial network. Translated in 2020 HEVC codec efficiency, the bitrate will be 65 Mbits/s. If the bitrate could be decreased by half (~33 Mbits/s), the business model for satellite would be reasonable, as one channel could fit on a legacy transponder, cable DOCSIS 3.1, or fiber infrastructure that could also support that model. If divided by 4 (~17 Mbits/s), a UHD channel could fit on a terrestrial network, multiplexed with other HD channels.

Looking at VR 3DoF in 2020, it is realistic to have a full UHD-1 120 (3840 × 2160 × 120) resolution as described in **Table 2**, which leads to a bitrate of 13 Mbits/s when using optimized VR transmission systems like tiling.⁴ If a new codec was used, a 50% bitrate reduction could be achieved at about 7 Mbits/s. A more advanced compression of 75% versus HEVC would bring the bitrate to below 3.5 Mbits/s. Both of these are attractive on 4G LTE networks.

VR

Looking at VR 6DoF in 2020, based on the information communicated by Intel at its Consumer Electronics Show 2017 press conference, an uncompressed frame will be 3 Gbytes, meaning 180 Gbits/s for video. This is based on point cloud techniques, for which compression is not yet defined, but may be within the scope of JVET. An uncompressed video is at 60 frames/s, 1500 Gbits/s. Assuming similar compression efficiency as HEVC, it may be possible to lower the bitrate to within the gigabits per second realm, which makes this application more suitable for storage media/download, and not streaming.

Mobile

OTT on mobile devices is a difficult situation as the network cannot sustain high throughput at present. In

the U.S., T-Mobile Binge On, whose concept is now being copied by Sprint with its Unlimited Freedom service and AT&T with its Stream Saver service,⁵ caps the video transmitted to the device inside the network to 1.5 Mbits/s for a max 480p resolution (using the Advanced Video Codec (AVC) codec). This is achieved using a signature in the encoded stream that, through a simple deep packet inspection scheme, filters the manifest (even if Digital Rights Management is applied). This shows that the mobile video traffic can scale on existing networks, of course, lowering the user experience.

Another promising technology is LTE broadcast. Here, the problem is different, as the video is transmitted at a single bitrate via multicast over a 4G mobile network. If by 2020 most of the smartphones support UHD resolution (3840 × 2160), then a minimum 1080p60 transmission will be required. Using HEVC, this corresponds to roughly 2.5 Mbits/s, which could be sustained on a 4G LTE network. If UHD-1 has to be transmitted in HEVC, it requires 13 Mbits/s. A new codec will definitely be needed for 4G LTE or LTE broadcast to bring down the bitrate to 7 Mbits/s with a 50% compression improvement and 3.5 Mbits/s with a 75% compression versus HEVC.

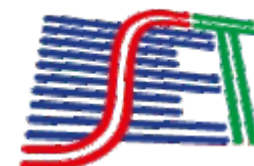
5G seems very promising, and it is conceivable to achieve 1 Gbits/s by 2020 for end users mobile devices, although the economics is not yet proven for those values. Staying conservative and only taking 100 Mbits/s is enough for all identified use cases 15 shown in **Table 2**. Only use case 6 requires the full 5G capacity, and in this case, it may be advantageous to work at the system level to reduce the bandwidth.

Download

In the initial days of internet video, download and progressive download were widely used. Later, this was supplanted by adaptive streaming. Still, adaptive streaming works only if there is enough bandwidth capacity. Recently, Netflix⁶ and YouTube, who have always stated that they would only stream content, have started to bend the rules. It began in emerging countries where bandwidth is not sufficient, but now it is

TABLE 2. Bitrate for next-generation 2020 applications.

Use case	Format	Application	Transmitted resolution	Bit depth	Codec	Bitrate (Mbits/s)
1	VR Tiling HD	Streaming	1920 × 1080 × 60	8 bit	HEVC	3.3
2	VR Tiling UHD	Streaming	2560 × 1440 × 60	8 bit	HEVC	9.8
3	VR Tiling UHD120	Streaming	3840 × 2160 × 120	8 bit	HEVC	13.1
4	UHD-120	Streaming/broadcast	3840 × 2160 × 120	10 bit	HEVC	19.7
5	UHD-2 (8K)	Broadcast	7680 × 4320 × 120	10 bit	HEVC	65.6
6	VR 6DoF	Download	Undefined	8 bit	HEVC	>1 Gbit/s



happening everywhere as good network coverage is hard to attain while commuting. This clearly shows that with better compression, streaming is possible. Of course, if the resolution keeps increasing, those higher resolutions might not be streamed. On the VR side, the move from 3DoF to 6DoF forces a download model, until the network copes with gigabit traffic.

New Compression Techniques Exploration

This section presents all of the different techniques that can make MPEG a superior standard versus any other solutions. All of the bitrate projections are indications based on either real product performances in a previous codec, proof of concepts, or pure research. The techniques presented here were introduced at the MPEG 114 meeting in San Diego in February 2016.⁷

MPEG JVET Target

Using traditional compression techniques, MPEG has initialized requirements for the future video codec with an objective of 50% bandwidth reduction, compared with HEVC at the same level of video quality by 2020. Early experiments at the end of 2016 showed that JVET can achieve a gain of 20% on HEVC. Therefore, it is expected that a 50% target is within reach by 2020. Meanwhile, it is safe to assume that in the beginning, HEVC efficiency on encoder products may not reach more than 35% gains compared with HEVC for live. It is not prudent to delve any further into the JVET topic, as it could be the subject of an entire paper.

Content-Aware Encoding

Netflix announced in 2015 that content-aware encoding can save 20% bandwidth versus classical fixed bitrate encoding.⁸ This approach basically uses variable bitrate per title encoding to provide more bits to elements of the titles where the content is more complex. Recently, Netflix refined its approach where the decision was not made at the tile level but at the chunk level.⁹ This technique is also applied to VOD by YouTube. Harmonic has been working on similar techniques but with a much lower granularity (i.e., frame level) and running in realtime, for the time being with the AVC codec, but the technique can also be applied to any other codec. A technical guide¹⁰ provides an overview of the technology used and how this can achieve compression gains versus constant bitrate on average. The gain is content dependent as opposed to a new codec like HEVC. Compared to a pure capped VBR scheme, savings represent 20%. The technology is, in general, transparent to the client as a standards-based codec is used. Some changes may be needed, as certain clients are managing the bitrate changes since the bitrate per chunk is not constant anymore. This technique can be applied to OTT, whereby chunks can have a variable size. It can also be used with classical statistical multiplexing, which is used today to power all broadcast delivery, cable, direct-to-home (DTH), and terrestrial transmissions.

Figure 2 shows how the savings are applied in relation to the function of the content complexity. This technique is applied to compression technologies and requires a substantial increase in terms of computation power to accurately measure the image complexity.

Elastic Encoding

The concept of elastic encoding involves using more CPU resources for complex content in order to compress video without impacting the video quality. On the other hand, when the content is simple, fewer CPU resources are needed to deliver an acceptable video quality. Over a long enough period of time (e.g., one day), it is expected that using this type of mechanism will lower the bitrate compared with the classical static approach where the same computation is used all of the time. Moreover, this technique is expected to use less compute resources over the same period of time in an average versus a static case. This approach is still in the research stage, and it is impossible to provide measurement results. One comparison involves halving the CPU resources of an encoder in the so-called high-density mode, to increase the bitrate by 15%–20%. Reversing the reason - ing, a similar compression gain can be expected if more CPU resources are used at peak time.

This technique can be applied to any codec. To take full advantage of the technique, it is necessary to add a new unconstrained mode to JVET where all possible tools could be used when maximum CPU is available on the encoder side. Of course, it is important to guarantee a standard decoder can process it in realtime. Note: The higher complexity resulting on the decoder side could be at limited resolution where more CPU resources can be made available. Looking back at the 1080p60 example, it is likely to expect a UHD-1 decoder to reallocate its decoding resources to decode more complex algorithms at lower resolutions. For the time being, a conservative approach is best. Here, one can assume that using HEVC elastic encoding could provide a 15% gain (as HEVC was not designed for

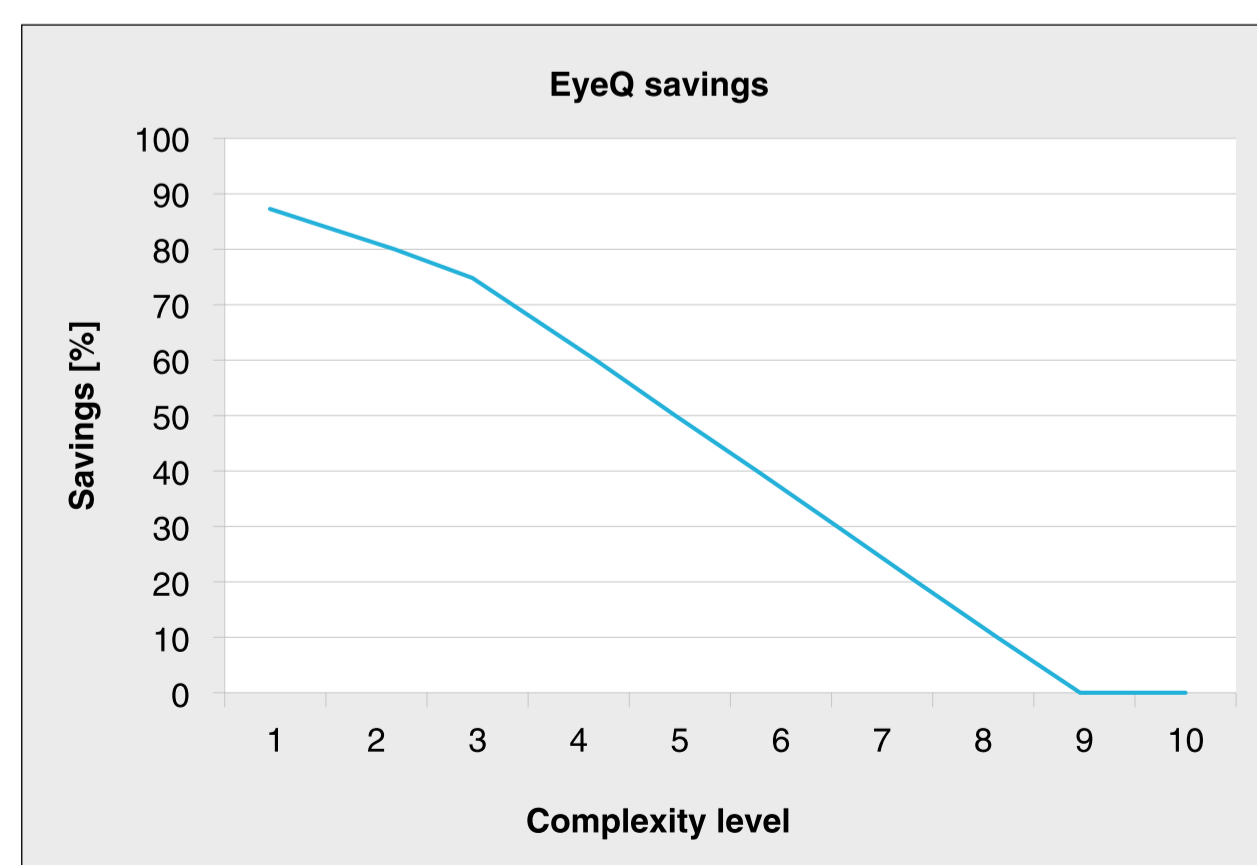


FIGURE 2. EyeQ chart.



that), and that with a JVET standard (assuming it is designed as elastic encoding-friendly), a 30% gain is possible. Those assumptions are projections and will have to be verified through simulations within the MPEG framework.

Machine Learning

Machine learning is becoming a hot topic in video compression, although it started 20 years ago with neural networks. More recently, with the advances of artificial intelligence and the availability of more compute resources to run exhaustive models, there has been a resurgence of machine learning for video. At this time, there are no known services that use this technology, but the recent acquisition of Magic Pony by Twitter is proof that this technique is becoming more mature.¹¹ The concept is to train the encoder offline and apply the rules learned at runtime. Of course, the encoder can also learn at runtime, but that requires a lot of additional inline resources, so the right tradeoff between offline and inline training has to be found. In a cloud environment, it is possible that during the night when no difficult content is encoded, the machine learning system uses available CPU capacity to learn daytime critical sequences. The performance of a machine learning system highly depends on the data set it has been trained on. For the time being, it is best to take a conservative approach and assume that with HEVC machine learning could provide a 15% gain (as HEVC was not designed for that), and that with a JVET standard (assuming it is designed as machine-learning-friendly), a 30% gain is possible. This technique can be applied to any codec. To take full advantage of it, MPEG should define reference test patterns and also provide reference software to show how machine learning can be used. Those assumptions are projections and will have to be verified through simulations within the MPEG framework.

Pre-/Post-Processing Pairing

Post-processing techniques have been developed on the decoder side for many years and shown improvements. A 30% savings can be achieved when post-processing is applied on the decoder side. This is based on very advanced filtering techniques that are applied after the decoder process and that reconstruct damaged video after heavy compression. This was jointly demonstrated on embedded devices, such as tablets and set-top boxes, at the International Broadcast Conference 2014 with encouraging results. We believe that if the encoder also applies some processing and metadata are transmitted to the post-process via supplemental enhancement information messages, this type of performance could be guaranteed across all types of content.

For the time being, a conservative approach is the best. Without pre-processing, the post-processor can only save an estimated 15%, with pre-/post-processing

pairing potentially reaching 30%. These assumptions are based on using the AVC codec, and they can be applied to any codec.

Video Quality Measurements

Picture-signal-to-noise-ratio was used in the early MPEG days to measure the performance of video compression algorithms. It is a simple method, but the video community does not believe that it is the best way to assess the video quality. We have moved to different video quality measurement techniques such as Structural Similarity Metric, Difference Mean Opinion Score, and Just Noticeable Difference. More recently, Netflix has introduced its own metric Video Multimethod Assessment Fusion (VMAF) that is based on a hybrid measurement of different metrics and that also includes machine learning. MPEG has not yet accepted VMAF as a standard way to measure the video quality, which is the main drawback of VMAF as it cannot be implemented in realtime due to its complexity. Harmonic, during the development of its EyeQ technology, has developed its own video quality metric that can be applied in realtime.

Bandwidth Reduction Roadmap

This paper has presented different techniques that can be used to go beyond pure codec compression. The next step involves exploring how HEVC can improve additional techniques, as just described, and use those techniques based on the next JVET standard.

With regard to improvements, it is expected that staying with the HEVC codec provides compression efficiency gains of about 10% per year. Comparing HEVC bare bones in 2020 versus HEVC boosted in 2020, the ratio is 2.2; whereas, looking at HEVC bare bones in 2016 versus HEVC boosted in 2020, the ratio is 3.3. **Table 3** provides a comparison of different HEVC deployment options.

In conclusion, if HEVC can be massively deployed, then the proposed improvements could bring a second life to HEVC and offer a 2× improvement in 2020 versus the HEVC bare-bones codec. As there are already approximately 2 billion HEVC-capable devices,¹² this could turn out to be a very attractive scenario.

TABLE 3. HEVC improvements table.

Bandwidth reduction	2016 HEVC	2020 HEVC (%)	2020 Boosted HEVC (%)
Core codec	NA	34	34
Context-based encoding			20
Elastic encoding			15
Machine learning			15
Post proc only			15
Total savings		34	70



Contributions	2016 HEVC	2020 HEVC (%)	2020 JVET (%)
Core codec	NA	34	35
Context-based encoding			25
Elastic encoding			30
Machine learning			30
Pre/Post proc			30
Total savings		34	83

Moving to JVET, the compression efficiency gain in 2020 will likely be only 35% versus HEVC (Table 4). This takes into account that, during the first year, the codec is not at full performance for live, commercial products.

Comparing HEVC bare bones in 2020 versus JVET in 2020, the ratio is 3.9. With HEVC bare bones in 2016 versus JVET in 2020, the ratio is 6. Figure 3 summarizes all of the different codec performances. With the proposed improvements to JVET by 2020, the solution comes close to 4x versus HEVC in 2020. This will, of course, have to be proven in simulation for the MPEG call for evidence that took place in October 2017.

Conclusion

Video compression is still a very relevant subject, especially if the industry wants to address new applications, mobile networks, and emerging countries. This paper described how new network-based topologies can bring massive improvements beyond pure codec compression, not only for new codecs but also for the existing HEVC codec. The different figures provide guidelines

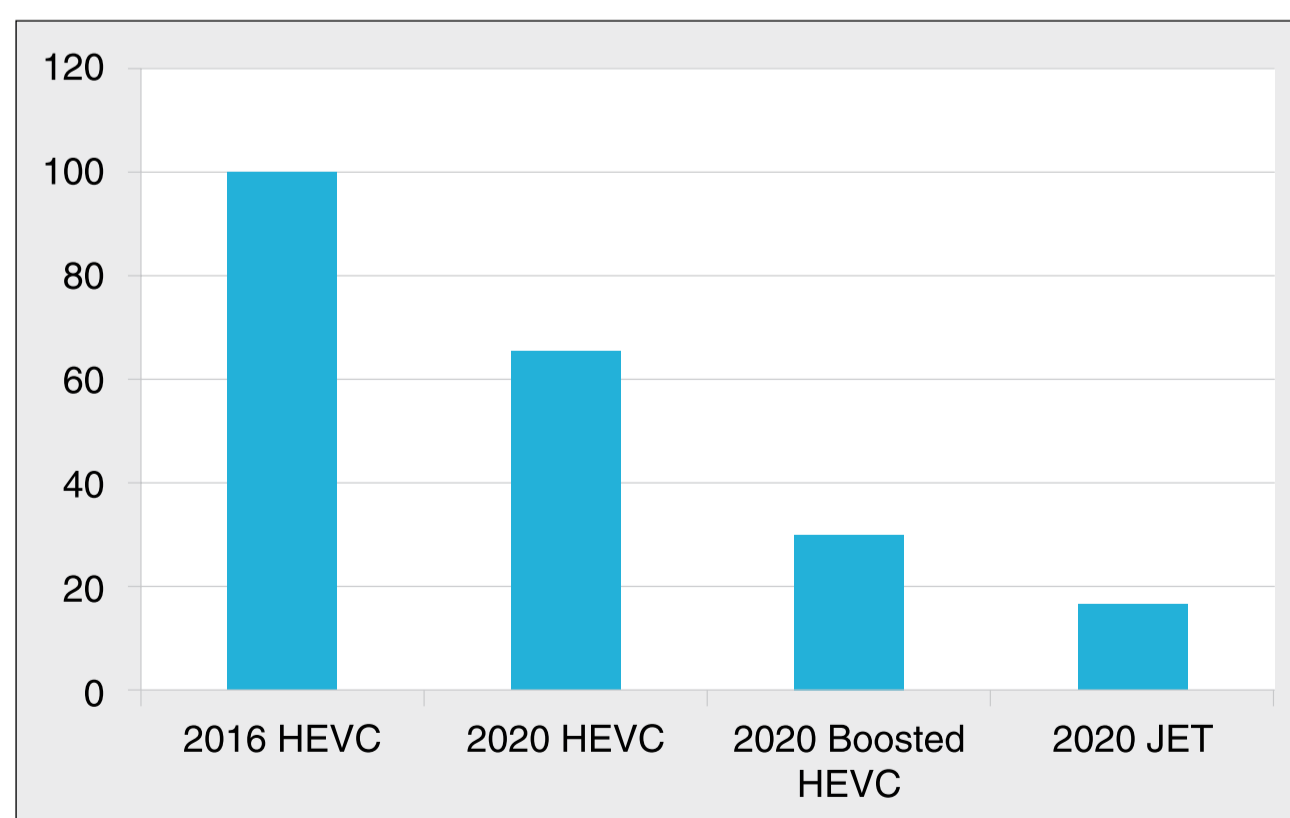


FIGURE 3. Comparison of codec performances.

for future work inside MPEG, either using HEVC or for the development of JVET. In either case, significant savings can be realized by taking on a holistic view of compression, which is now possible with new technologies available in the headend, in the network, and in the device.

Acknowledgments

The author would like to thank Leonardo Chiariglione, MPEG convener, who has encouraged him to present out-of-the-box thinking to MPEG for the next-generation video codec as well as Xavier Ducloux who co-chairs the Network Distributed Video Coding Ad Hoc Group in MPEG.

References

1. T. Fautier, Opinion: Is Codec War in Our Future, Streaming Media, Aug. 2016. [Online]. Available: <http://www.streamingmedia.com/Articles/Editorial/Featured-Articles/Opinion-Is-A-Codec-War-in-Our-Future-112919.aspx>
2. Requirements for a Future Video Coding Standard v3, The Moving Picture Experts Groups MPEG 115 Geneva Meeting, May 2016. [Online]. Available: <http://mpeg.chiariglione.org/standards/exploration/future-video-coding/requirements-a-future-video-coding-standard-v3>
3. MP20 Workshop on MPEG Roadmap, The Moving Picture Experts Groups Chengdu, China Meeting, Oct. 19, 2016. [Online]. Available: <http://mpeg.chiariglione.org/news/mp20-workshop-mpeg-roadmap>
4. Harmonic to Showcase First Public Virtual Reality Demowith Native UHD Resolution at IBC2016, Aug. 2016. [Online]. Available: <https://www.harmonicinc.com/news-events/press-releases/read/harmonic-showcase-first-public-virtual-reality-demo-native-uhd-resolution-ibc2016/>
5. AT&T to Optimize Video with Stream Saver, WhistleOut, Nov. 2016. [Online]. Available: <https://www.whistleout.com/CellPhones/News/att-to-optimize-video-with-stream-saver>
6. Finally! You Can Download Shows from Netflix, USA Today, Nov. 2016. [Online]. Available: <http://www.usatoday.com/story/tech/news/2016/11/30/netflix-finally-adds-offline-mode/94659494/>
7. Fautier and Fogg, Commercial Requirements for Next-Generation Video, MPEG/M37665, Feb. 2016.
8. A. Norkin et al. Per-Title Encode Optimization, Netflix Technology Blog, Dec. 2015. [Online]. Available: <http://techblog.netflix.com/2015/12/per-title-encode-optimization.html>
9. More Efficient Mobile Encodes for Netflix Downloads, Netflix Technology Blog, Dec. 2016. [Online]. Available: <http://techblog.netflix.com/2016/12/more-efficient-mobile-encodes-for.html>
10. EyeQ Achieving Superior Viewing Experience, Harmonic Tech Guide. [Online]. Available: <http://info.harmonicinc.com/Tech-Guide-Harmonic-EyeQ>
11. J. Dorsey, Increasing Our Investment in Machine Learning, Jun. 2016. [Online]. Available: <https://blog.twitter.com/2016/increasing-our-investment-in-machine-learning>
12. HEVC: Are We There Yet? Frost & Sullivan, Jul. 2016. [Online]. Available: http://conferences.infoday.com/documents/257/A101_Rambhia.pdf



About the Author



Thierry Fautier is vice president of video strategy at Harmonic, San Jose, California, where he is in charge of defining and driving the execution of the long-term strategy of Harmonic video business. In addition, he is currently president of the Ultra-HD Forum, the global organization responsible for promoting market adoption of UHD by defining industry best practices for the phased introduction of the wide set of technologies that facilitate the next-generation viewing experience. His previous experience at Harmonic includes leadership positions

in solutions marketing, where his area of responsibility covered Harmonic major markets, including broadcast, cable, telco, DTH, and OTT, as well as multiscreen initiatives. Prior to joining Harmonic, he was the vice president of marketing for Vsofts and held various positions at Philips in research, engineering, and marketing. He was one of the primary drivers at Philips in the creation of Motion Picture Experts Group (MPEG) technology and in the development of the first MPEG-2 set-top box system-on-chip solutions. He is an experienced speaker and has presented at leading industry conferences across the world.

FOLLOW US



smpteconnect

