Driving codec efficiency

Breaking down the bits

MediaKind | Application Paper

Presently, there are more competing codecs than at any time before. The existing standards are still in use in many places, yet new codecs are vying for attention as the ideal answer to media content delivery. Often, these are presented as a simple comparison “A is better than B”, but frequently the situation is vastly more complex and requires more context. In this paper, we’ll take a look at the latest candidates, provide an assessment of characteristics, and identify where they might fit in the various media ecosystems.

The following codecs are included in this white paper:

- High Efficiency Video Coding (HEVC)
- Essential Video Coding (EVC, Baseline and Main Profiles)
- Versatile Video Coding (VVC)
- AOMedia Video 1 (AV1)
- JPEG-XS

Codecs and encoders

First, a quick digression. There is a difference between codecs and encoders. Codecs, which really ought to be known as Coding Standards, define the bitstream syntax and required operation of the decoder. This is necessary because it is important that every decoder produces an identical output for a given input bitstream. Encoders, on the other hand, are specific implementations of a codec. In most cases, an encoder is free to choose any way to implement the codec, as long as the output (the bitstream syntax and in many cases the timing of the bitstream) is legal (i.e., is compliant to the coding standard). Over time, increasing expertise and compute power mean that more efficient encoders can be built, while retaining compatibility with existing decoders.

This brings us to the first consideration when trying to evaluate a codec: we are in fact comparing encoders (i.e., specific implementations of the encoding end of the codec), so the encoders may or may not reflect the performance possible with a given codec in another implementation. Does it matter? Absolutely it does. The variation between different implementations in terms of compression efficiency can easily be much larger than between codec, so it is important to consider the actual implementations used, along with their configurations.

Reference encoders

While codecs have a normative decoder (explicitly and specifically defined in the coding standard), most codecs also have a reference encoder implementation (an instantiation of an encoder created for comparison or verification testing) . In recent years, the reference encoders have tended to be bitrate efficient but slow, aiming to illustrate the potential compression efficiency of the codec. The degree to which this is true varies between different codecs’ reference encoders, with other variables such as the presence and type of rate control, or the use of tools such as adaptive quantization, making a significant difference between the reference encoder and the assumed asymptote of performance (assumed, because over many years we have seen encoder algorithm innovations that have driven encoder performance for some existing codecs significantly beyond what was assumed to be possible). With such

codecs, there are usually commercially produced conformance bitstreams that explore the corner cases and limits of the standard, ensuring that the decoders really do support the standard fully.

The arrangement of normative decoders and encoder freedom means that commercial encoder vendors can benefit from investing in making those efficiency improvements, with competitive bitrates typically reducing by about 10% per annum while maintaining the same image or picture quality. Ultimately, this is of benefit to both the encoder vendors and the customer base, since it reduces the cost of delivery, whether that is satellite, cable, IPTV, terrestrial or adaptive bit rate.

Library-based codecs

A few codecs provide implemented encoder libraries, rather than reference encoders. These offer a ready-made encoder implementation that can be built and run with minimal effort. Clearly, this reduces the barrier to entry for creating encoders. In some cases, these libraries are open-source, meaning that contributions can be made to the libraries, improving the efficiency from either a bitrate or a compute requirement perspective. Once the contribution has been included, then it is available to all users of the library.

On one hand, this provides a reasonable implementation for the encoder; on the other hand, however, it removes the incentive for commercial encoder vendors to invest in efficiency improvements, since they would either be available to everyone, if contributed to the library, or would create a recurring cost for integration of the changes with every update to the library.

Another potential issue that acts as a barrier to commercial investment is that the responsibility when an interoperability issue occurs is somewhat unclear. Assuming there is not a syntactic error, there can be ambiguity in where the fault lies: if the behavior is legal but significantly different from how the library implementations behave, is that still expected to be supported by all decoders? Are there conformance bitstreams to ensure that all legal behavior is supported?

Evaluation criteria

Evaluation of codecs is a difficult area, primarily because the real aim is to be able to measure bitrate (which is easy) for a given visual quality as experienced by a human viewer (which is extremely difficult). There are some easily created metrics, such as peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), which have the benefit of being well understood – however, both suffer from being overly simplistic and can diverge materially from the perceived image quality. They are, however, repeatable and tend to be used by some customers as satisfactory proxy for visual quality. More complex, and better correlated, metrics can be made, but there is an absence of standardization, which limits the desire to use them. More recently, some organizations have used machine-learning techniques to weight a basket of different, simple metrics based on content features assessed by the neural network. While this does offer some ability to bias away from the more misleading cases, it cannot itself create measurements of visual artifacts that are not already included in the simple metrics – for example, if all the metrics are spatial-only, then no amount of learning how to weight them will represent temporal artifacts being represented.

In this paper, we will use PSNR and SSIM as our metrics for comparing visual quality. As above, some caution is needed when comparing different codecs, since if, for example, one produces softness (a coding artifact that manifests itself as a visible reduction in localized image resolution) and the other produces ringing (a coding artifact that manifests itself as a halo around object edges) when pushed hard, then the results will be reflected differently in the different metrics, even though they might be considered “comparably impaired” by a viewer.

A final note relates to the measuring of adaptive bitrate (ABR) bitstreams. These have an additional complexity, since ideally we would like a single figure that represents the performance for the full set of representations. However, each representation will have its own score and it is not immediately obvious that there is an ideal means to compare them. Summing the bitrates for all representations would bias towards the highest representation, whereas averaging percent savings would favor making the lower representations better.

With the ever-increasing screen resolution of the latest generations of handheld devices, it would seem reasonable to consider the highest representations as being the most important, given that the expectation of viewing quality will be highest for those cases, and because those high representations contribute more to the bitrate required, and hence cost.

Additionally, when considering codecs, the complexity (i.e., the compute resources required) for both encode and decode need to be considered. Depending on the use case, these will have different levels of importance: for example, user-generated content might be expected to consider the encode cost as more important than bitrate efficiency.

2020 codecs

The following sections are a review of the current status of each of the main codecs in 2020.

AVC (H.264)

Advanced Video Coding (AVC) was standardized in 2003, so can be considered mature and encoder implementations are well optimized. Approximately 10 years after the venerable MPEG-2 Video standard (which changed the media and entertainment industry forever) was published in 1994, AVC's goal was to achieve a coding efficiency of 50% less bitrate than MPEG-2 Video for the same perceived image quality. The timing of AVC's availability matched the rapid deployment phase of High Definition (HD) TV around the world, and was an obvious choice for all but a few very early systems. Despite being nearly 20 years old, AVC is still the dominant codec in use for both HD TV and for ABR streaming. AVC encoders are still being made more efficient today.

HEVC (H.265)

High Efficiency Video Coding (HEVC) was standardized in 2013, ten years after AVC, with a target of achieving a coding efficiency of 50% less bitrate than AVC for the same image quality. Unlike with AVC, the migration to Ultra High Definition (UHD) did not immediately accelerate at pace. To some extent, this was because around the same time, the desire to make the video experience better, rather than just more pixels, occurred, driven by a combination of High Dynamic Range (HDR), Wide Color Gamut (WCG) and also a general recognition that 8 bits per pixel were not enough – causing unpleasant banding that simply didn't match the expectation of UHD's viewing experience. With the adoption of Main 10 Profile, and despite many disagreements on HDR systems, HEVC emerged as the appropriate standard for all UHD TV deployments. However, as the delivery options increased, and to some extent the uncertainties about HEVC licensing took hold, some of the UHD systems considered alternatives or were postponed. In the conventional TV world, HEVC remained the industry-wide choice, however for streaming platforms, there has been more variability. HEVC's test model is known as "HM".

AV1

AV1 was developed by the Alliance for Open Media (AOM) consortium, as a competitor to HEVC, with a similar efficiency target. In 2017, there was a "software freeze" for the reference code, followed by a formal specification in 2018.

The intent for AV1 was to create a format that was royalty-free. The reality of video compression is that there has been a substantial amount of research by many companies around the world, so ensuring that there is no existing intellectual property rights (IPR) that isn't from a consortium member is difficult to achieve. At the time of writing, Sisvel has a patent pool, with associated fees, for AV1 covering approximately 1400 patents that are claimed to be relevant to AV1. As a result, the initial aim of having a royalty-free codec does not appear to have been achieved.

There are two main threads of libraries developed:

- "LibAOM" – this is the reference encoder, which targeted high bitrate efficiency at the expense of processing time
- "SVT-AV1" – a library that started with the aim of minimizing compute required

Thus, there are two sets of figures for AV1, reflecting the different models. Both of these are complete libraries, so easy to adopt, but less room for differentiating implementations.

Over time, LibAOM has become faster to execute and SVT-AV1 has become more efficient, albeit at the expense of more compute resource. The two libraries are now very similar in their compression efficiency.

In addition to the encoder libraries, Dav1d is an open-source software decoder for AV1, which can be implemented on multiple processor types. Along with the software decoder, implementations for AV1 in both FPGA and silicon exist, so the potential for adoption is increasing.

VVC (likely to become H.266)

Versatile Video Coding (VVC) is the successor to HEVC in the standards path (a joint development from ISO/IEC MPEG and ITU-T VCEG). Once again, the ambition was to achieve around 50% bitrate saving compared to the previous standard, HEVC. VVC was released as a final standard in July 2020, with some uncertainty about the license terms.

VVC was designed to support HDR and WCG, and also includes native support for some new use cases, including:

- Screen content coding, aimed at optimization for gaming, computer screens, etc.
- Adaptive streaming with resolution switching based on Reference Picture Resampling
- Immersive media, with coding of independent sub-pictures for tiled streaming of 360-degree videos

VVC's reference encoder model is known as "VTM".

EVC (MPEG-5 part 1)

Essential Video Coding (EVC) is an attempt to run a fast-track standardization exercise, with a limited number of companies involved. EVC has achieved Final Draft International Standard, with standardization expected to complete during 2020. EVC is, in fact, two different standards: Baseline Profile and Main Profile. While the use of the term "profile" has historically been used in a manner that profiles build upon each other, in this case the two profiles of EVC are very different:

EVC Baseline Profile is a royalty-free codec exclusively using IPR that has expired patent terms.

EVC Main Profile is a royalty-paid codec that uses IPR from only a few companies, making the license terms easy to negotiate; however, it is impossible to preclude the possibility that some of the IPR included in EVC Main Profile may be subject to an IPR assertion, in a similar manner to Sisvel's pool for AV1. EVC Main Profile has a fallback position that if 3rd party IPR emerges, then there is a lower-performance alternative that could be used – albeit with a bitrate penalty. Each tool is isolated and may be switched-off independently of other tools.

Since the Baseline Profile is not aimed specifically at bitrate efficiency, it is not compared in this paper. Instead, Main Profile is evaluated using its reference model "ETM".

JPEG-XS

One further recently available codec is JPEG-XS. JPEG-XS is somewhat different from the other codecs reviewed here, as its target application is ultra-low latency near-lossless operation. JPEG-XS uses a small, but configurable, number of lines for the compression blocks, meaning that the latency can be kept down to a small number of lines. The consequence, though, is that JPEG-XS is much less efficient than the other codecs here. However, that does not dilute its value, since it is designed mainly for in-studio or across fiber connections during the production process, where the priority is to keep near perfect picture quality and ultra-low latency is important because of live switching and synchronization of sources.

The use of a relatively small number of lines in JPEG-XS's compression process is a significant difference from JPEG-2000, where the hierarchical picture structure of compression necessarily causes at least one picture delay at each end of the link. JPEG-XS's efficiency of around 10:1 compression makes the infrastructure needed for connecting studios via IP using SMPTE ST 2110 (real-time Professional Media over IP) much more cost-effective. More details can be found in our Remote/At Home Production application paper here: https://www.mediakind.com/insights

Performance comparisons

In general, the comparisons shown below are referenced to HEVC, to provide a consistent baseline. Despite its limitations, PSNR has been used as the comparison figure. As a result, it is important to use the information as a general understanding of the compression capabilities, and avoid reading too much into any relatively small differences.

For each codec, as the reference model progresses, new versions are released, potentially with significantly different characteristics, so the specific version of the test models (or library implementations) being compared are shown in the comparisons.

VVC

VVC's performance has changed significantly as the reference model progressed, as shown in Figure 1. With VTM-9.0, the efficiency approaches 40% compared to HEVC (HM-16.19), but this is averaged over content formats. For higher resolutions (e.g., UHD), the bitrate saving is more than 40%. Compared to the HEVC reference model, the encoder requires roughly 10x the processing power for the encoder and just under 2x for the decoder. While significant, this level of change is reasonable in the context of improvements to compute performance over the interval between the standards being finalized.

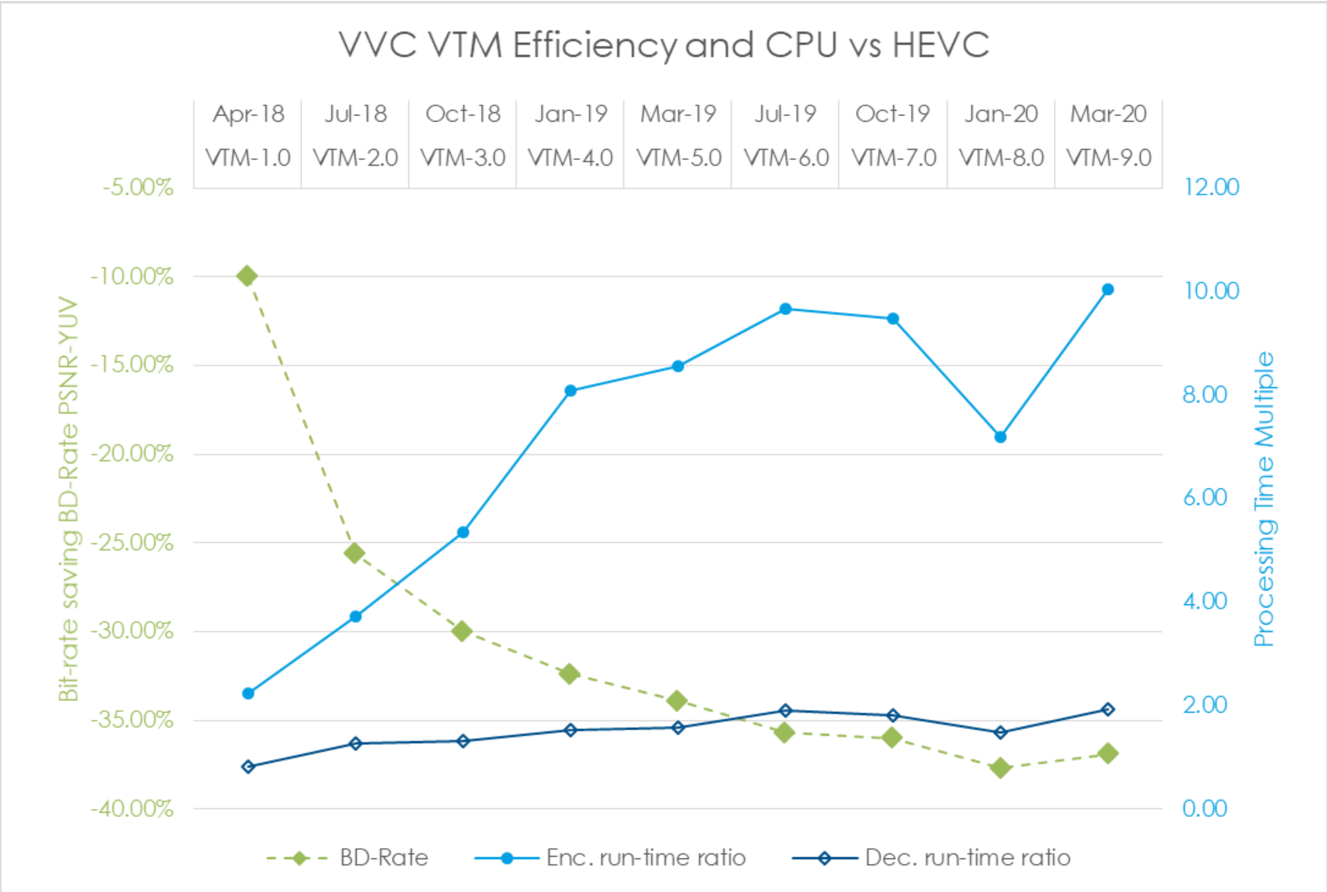

Figure 1: VVC efficiency compared to HEVC (HM-16.19) as the reference models progressed

VVC and EVC main profile

When comparing VVC VTM-9.0 to EVC Main Profile ETM-6.0 in Figure 2, it is clear that VVC is more efficient than EVC Main Profile. There is a lower compute requirement for EVC, compared to VVC, but still several times that required for HEVC. EVC can be viewed as largely a subset of VVC, because the structures are very similar. The consequences of this are that there is a potential for a significant level of commonality in EVC Main Profile and VVC in both encoder and decoder implementations, meaning an opportunity to migrate from one to the other.

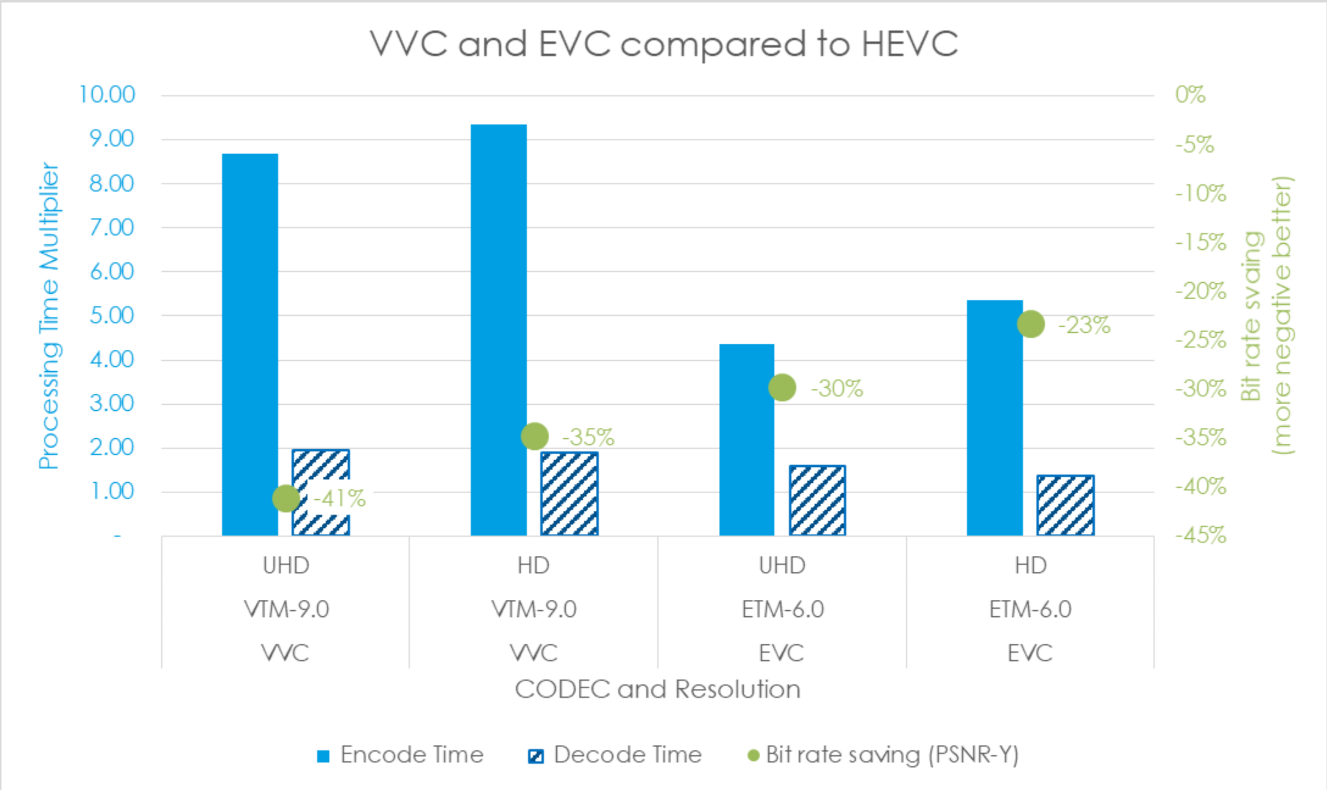

Figure 2: Comparison of VVC and EVC main profile for HD and 4K UHD

AV1

AV1's performance relative to HEVC has been a matter of significant debate over time. To some extent, this can be attributed to the choices made for evaluation and the metrics used. Subjectively, it appears that AV1 tends to lose detail with lower bitrates, which has a tendency to give relatively better results when measured using SSIM, but poorer results when measured with PSNR.

LibAOM has a number of different configurations, for different tradeoffs between processing power required and the resulting bitrate efficiency. For the tests involving AV1, the highest coding efficiency configuration was used.

AOM is also developing a future standard, AV2, the details of which are yet to emerge, but contributions are being incorporated into an experimental branch.

Overall comparison

Figure 3 shows the performance of VVC, EVC Main Profile, LibAOM-AV1 and SVT-AV1, indicating how their performance has progressed over time. From this it is clear that all of the candidates can out-perform HEVC, with VVC offering the biggest reduction in bitrate, followed by EVC and then AV1. When evaluated subjectively, the results are consistent with the order.

Interestingly, both implementations of AV1 have iterated to a similar level of bitrate efficiency.

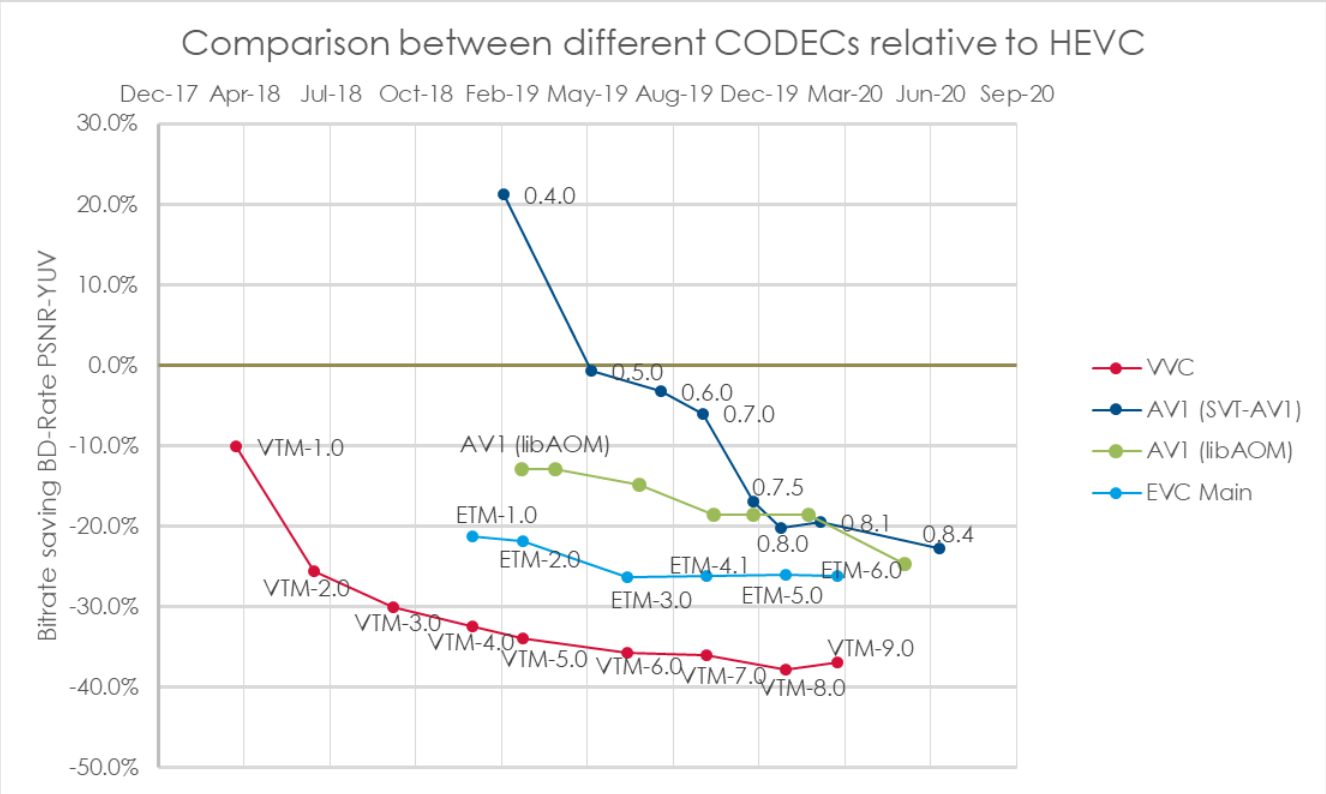

Figure 3: Comparison of new codecs relative to HEVC

Innovations

The performance of reference models for codecs is only one aspect of encode efficiency. As mentioned previously, real encoders are actual implementations of algorithms that typically differ from the reference models. As a result, even long after standardization, improvements can be made to encoder algorithms beyond the reference model's structure. This is the role that encoder vendors with research teams play, with improvements still being made to encoders for all standards, from MPEG-2 Video to HEVC – and VVC in the future – using knowledge of the human visual system's characteristics. Alongside the algorithm improvements, system-level improvements are also possible, such as MediaKind's Constant Video Quality (CVQ) for ABR, whereby the encoder avoids using more bits than necessary to achieve a required picture quality for easier scenes.

Mapping codecs to applications

In order for a codec to be successful, there needs to be an appropriate market application. For AVC, this was the arrival of HD, creating a requirement for new infrastructure and thus enabling a new codec. UHD and the continual need to be save bandwidth whilst enabling more and better entertainment services in linear TV is following a similar path and adopting HEVC, but what about other codecs and applications? AV1 was conceived as an efficient new codec that would be easy to adopt because of its royalty-free initial aim, which could outperform HEVC. While the efficiency has met its objective, the reality for licensing seems different and this likely will make its adoption more difficult. Undoubtedly, the complexity of HEVC's multiple patent pools has caused concern and provided a rationale for the use of AV1.

At the same time, EVC Main Profile has emerged after a very short standardization process and performs better than both HEVC and AV1, again with a relatively simple licensing scheme. This could make it a preferable choice for some cases. However, with the impetus of major organizations behind AV1, it seems likely that it will continue to be seen as the successor to VP9 in some cases, particularly for web browsers, social media and some streaming applications. This uncertainty means that media technology suppliers will likely need to provide multiple codecs, particularly in devices such as mobile handsets, meaning several codecs will remain in use concurrently. Regardless, it is possible to identify some subsets that might map well.

Pay and Free-to-Air TV

In general, Pay and Free-to-Air TV share a common constraint of client devices that are infrequently updated. This means that it is relatively difficult to migrate from established formats to new ones; hence, why MPEG-2 Video is still in use today. It is reasonable, however, to expect new services to be launched with the latest formats, so HEVC is the current choice for 4K UHD while later deployments could use VVC for better efficiency. 8K UHD formats are likely to adopt VVC, apart from a few early pilot services that might not have decoders available.

As STB population become closer to being all HD capable, except for a few places where MPEG-2 is used for HD, it becomes possible to turn off the MPEG-2 Video format and recover a significant capacity for other services.

ABR & Over-The-Top

For ABR, the picture is more complex and has a higher dependency on the type of content. While royalty-free codecs are always attractive, the licensing for both HEVC and AV1 are device-based, so it is likely that both will be present on many consumer devices. As such, it seems likely that the ABR market will use (in addition to its continued use of AVC) both HEVC and AV1, most likely roughly along the existing AVC/VP9 split, although some of the web-scale SVOD services may well exclusively use AV1 at higher resolutions.

Remote/at-home production

JPEG-XS is somewhat different in nature from the other codecs discussed in this paper and is a strong candidate for adoption in studio and remote/at-home production environments, where the need is for extremely low latency and near-perfect image quality. JPEG-XS can make IP-based studio and production systems viable, including allowing 4K UHD to fit comfortably in standard 10G and 25G Ethernet connections.

Virtual reality and live 360° video

Finally, the world of VR has a set of needs that extend the conventional formats – the need for parallel, but separable, streams of video for different areas of the picture is met by VVC's updated high level syntax. VR/360° Video, like 8K UHD, requires a significantly higher bandwidth than existing formats, so any tools that help to reduce the cost of delivery are important – in this respect, VVC is a significant enabler of both.

The future of codecs

The last few years have seen codecs released at an increasing pace, making choices for the media industry more difficult. Indeed, the list of codecs explored in this paper omits a number of more niche codecs, for example country-specific or single-vendor codecs. Future codecs for existing applications will need to be significantly more efficient than those already defined, in order to justify transitions. However, the trend of decode devices becoming more powerful will continue, as lower-power devices can more easily adopt the smaller geometries that enable increased processing power, which in turn enables more complex predictions, filtering, transforms and entropy decoding to be performed. In addition, the maturing of machine learning (ML) promises to allow new ways to form better predictions and to learn statistical patterns through training rather than being explicitly defined. However, in order to have a deterministic decoder, any client-side ML must behave identically on all implementation platforms – something that isn't typically the case today for different ML architectures. As future demands for video increase in quantity, resolution, fidelity and frame rate, the need for more efficient codecs is likely to continue. So it seems certain that new, innovative techniques will be developed leading to a further reduction in bit rates in the future.

Acquire
amazing.

Deliver
dynamic.

Experience
extraordinary.

mediakind.com