Concatenated encoding: Effects of mezzanine link compression performance

Overcoming limited bandwidth: How operators can utilize mezzanine links to efficiently impact QAM capacity

A Technical Paper prepared for the Society of Cable Telecommunications Engineers
By

Scott Davis
Principal Engineer, Video
Time Warner Cable
12101 Airport Way
Broomfield, CO 80021
+1-303-801-3005
scott.davis@twcable.com

Tony Jones
Head of Technology, Compression
Ericsson Television Ltd
Comines Way
Hedge End
Hampshire UK, SO30 4DA
+44-23-8057-3269
anthony.jones@ericsson.com

Overview

Cable bandwidth remains an extremely valuable commodity. While new compression technologies are being standardized, the cost of replacing deployed STBs (set-top-boxes) means that existing video compression standards will be in use for many years. Additionally, while recovering bandwidth is critical, the consumer experience cannot suffer as a result. This paper will explain how mezzanine links from content owner to operators can considerably improve the efficiency of utilization of QAM capacity, compared to the existing relatively high compressed MPEG-2 source feeds often available by satellite and provide a high quality end user experience

The paper will explore the effects of the concatenation of differing coding standards, formats and frame rates leading to candidate operating points. Building upon Ericsson's twenty years of video compression technology leadership, the company will be joined by a major US cable operator to provide an insightful look into the technologies that will enable this real world change today and tomorrow, discussing how operators can maximize the revenue opportunity represented by the introduction of mezzanine links, combined with the next generation of TV compression technology. The joint session will describe how the results demonstrate improved efficiency of use of QAM resources.

Content

1 Motivation

1.1 Consumption of Available Bandwidth

Pressure on available bandwidth continues to outpace any increase in capacity. In particular, the continuing roll-out of HD services, increasing popularity of on-demand content as well as DOCSIS data use for both general internet and for video on tablets, laptops, etc. together cause a need to make best use of the precious commodity that is bandwidth to the home.

1.2 Factors Affecting Bandwidth Use Efficiency

Several techniques, such as better encoders, statistically multiplexed groups or SDV have been used to improve QAM utilization efficiency; but there is considerable value in further improving the utilization efficiency. While this analysis has been focused around cable, the conclusions apply more generally.

Effective bandwidth utilization can be improved by using more than one of these techniques. One area for specific attention is the effect that concatenated encoding, in other words more than one stage of compression, has on the efficiency of bandwidth utilization for the final emission encoding to the home.

It is also important to recognize that the ability to deliver content at a desired quality within a particular bit rate is very much dependent on the content, so it is the most challenging content and operating points that will determine the overall system capacity.

2 Effects of Content Type

As is well known, compression efficiency for any given encoder varies based on the content complexity and on the operating point. In most television systems, the objective is to achieve the lowest bit rate allocation that allows the vast majority of all content to meet the quality requirement and particularly so for the content that is considered the most valuable. If we take a look at a single encoder, starting at baseband video, then there is a probability distribution of needing a certain bit rate to meet a target quality.

For the sake of clarity, this paper will ignore the definition of “quality” in the early sections; however this is covered later for completeness.

In general terms, content complexity varies as shown below:

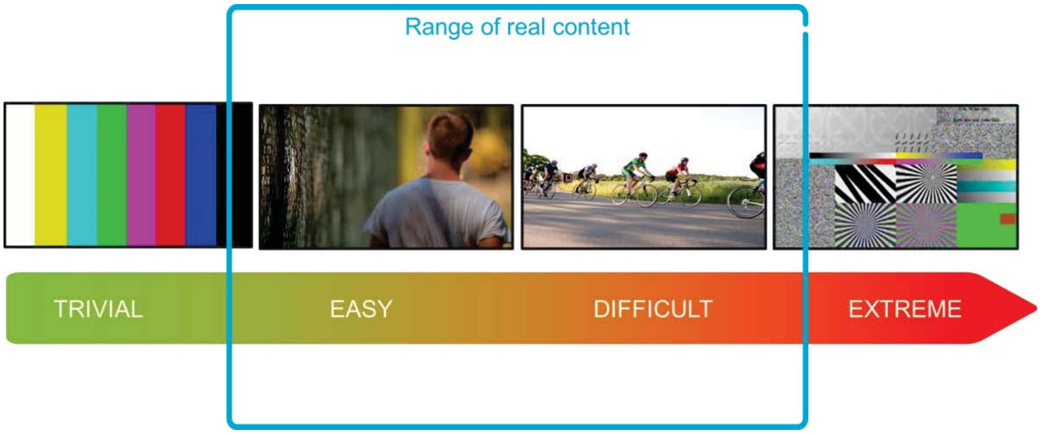


Figure 1: Variation of content by category

Between the extremes of synthetic test content lies real content, with (in general) movie type content being relatively low complexity and sport being high complexity. It is the high complexity content that will ultimately limit the performance of the system, so merits the most attention.

However, all content does not appear with equal probability, so we can draw a graph of the likelihood of different content complexity (and therefore bit rate demand) for a range of content.

Most real-world systems are not designed to ensure that absolutely all content meets a desired quality, since that would effectively lead to an excessive bit rate allocation to accommodate the long tail, but instead are configured to operate at a bit rate that makes the quality for the vast majority of real content acceptable to the home viewer. In particular, attention must be paid to the types of content for which the viewer will be most critical. Notably sports events have a high value to consumers and advertisers alike, so given that sport tends to be at the more complex end of the spectrum, it is a good candidate for the focus for performance.

If we were to plot the likelihood of needing a particular bit rate to meet a target quality, then its dependency on the type of content would be similar to Figure 2 below:

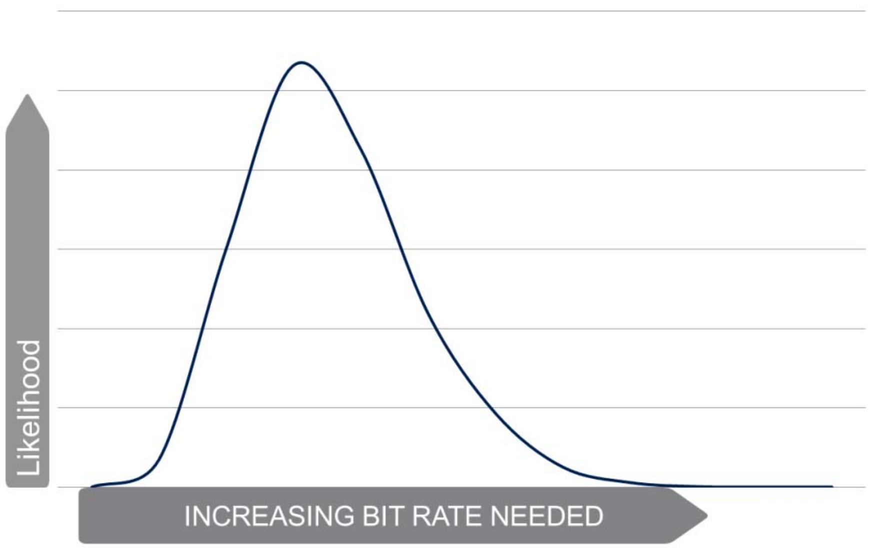


Figure 2: Probability of requiring bit rates to meet quality target for varying content

If the channel delivering the content has a restricted capacity, then content that lies above that limit will not meet the quality target. This is shown as the shaded area in Figure 3. The red line represents the channel bit rate limit and the content to the left of that line meets or exceeds the desired quality.

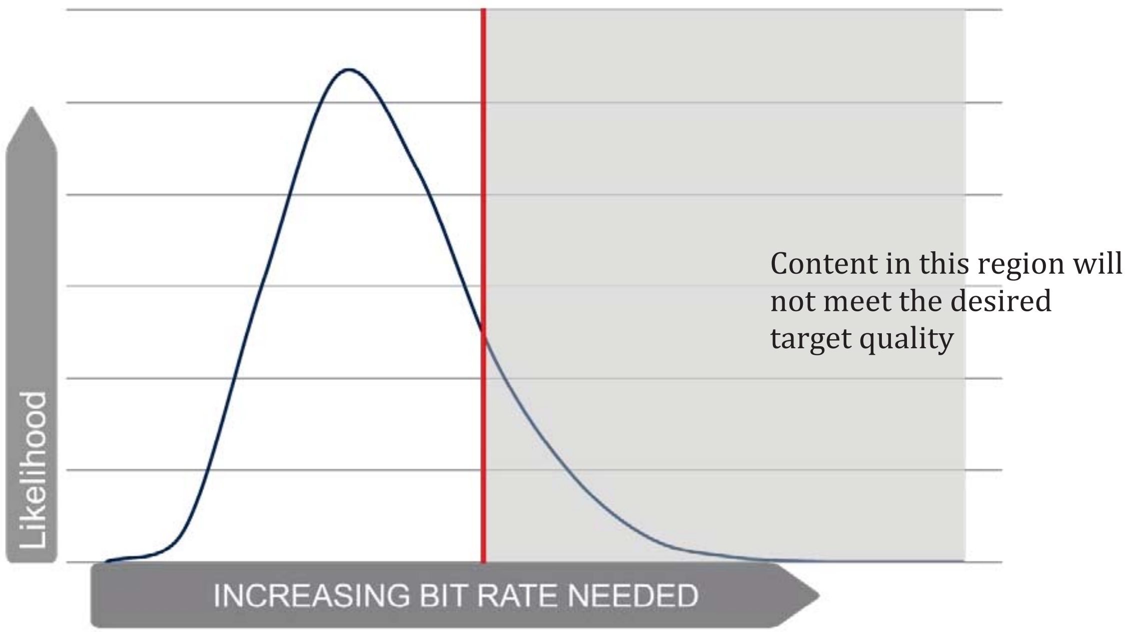


Figure 3: Content failing to meet target quality based on a fixed bit rate

Figure 4 below illustrates the difference in complexity between the clips using HD-SDI as the source.

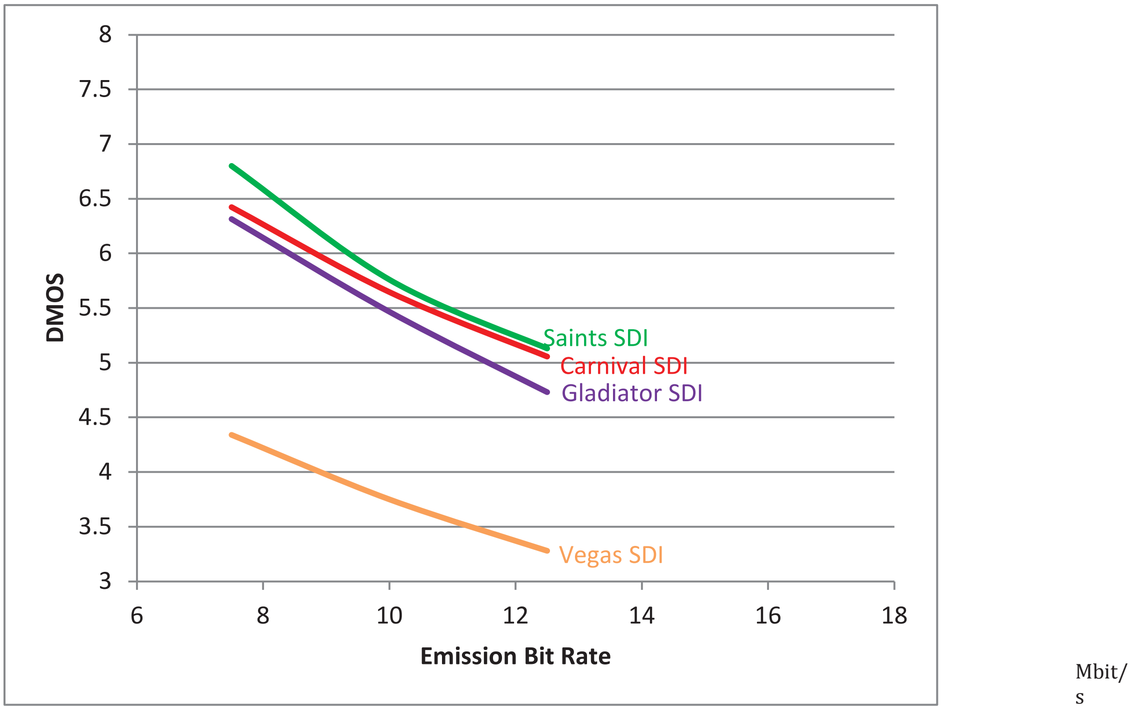


Figure 4: Complexity comparison of different clips

The above chart uses a measurement known as DMOS (Differential Mean Opinion Score - see later for details). DMOS indicates the level of impairment, so lower DMOS scores reflect higher quality, we can see that Saints (sports content) is significantly more difficult to encode that Vegas (easy movie content). In fact, a ratio of 2:1 in bit rate for the same quality is not unusual.

This supports the distribution shown in Figure 1 and implicitly informs that evaluation of performance with complex content (such as sports) will yield a result that will give good performance to all other content too.

3 Performance Characteristics

Taking the information from Figure 3, we can see that in order to optimize the efficiency, we need to pay more attention to content that is at or near the region where performance expectations are not met. We can see the reason for this by comparing the two quality vs bit rate curves below, which are for two specific pieces of content, one easy, the other more challenging.

The performance curves are characterized for a current, high-performance, encoder running MPEG-2 HD in a typical configuration for North American Cable, set to constant bit rate to allow comparisons to be clearly made, however the results apply equally to other resolutions and encoding standards.

The system is a single encode stage as shown below, representing the situation if the emission encoder were encoding from baseband.

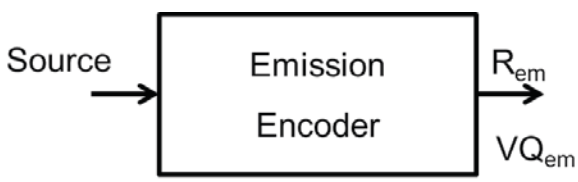
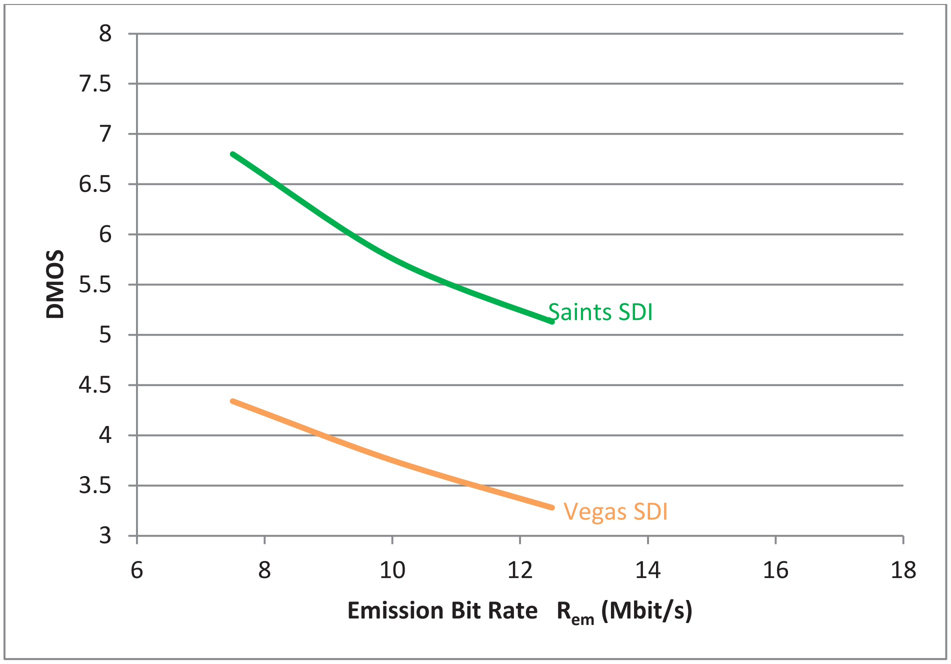


Figure 5: Reference encoder performance testing system



Mbit/
s

Figure 6: Quality vs bit rate for easy and complex content encoded from HD-SDI

It is reasonable to expect, therefore, that the effect of successively encoding each type of content will likewise vary, so examining the performance with concatenated encoding in the two cases should reveal information that allows an understanding of where changes are likely to be most beneficial

Clearly, encoding in a previous stage before the final emission encoder has the potential to impair the final emission quality: the question is the degree of impairment, and how that may be improved.

4 Concatenated Encoding

Most sources of content are today encoded several times prior to the final emission encoding. There are several reasons for this, but the primary reason has typically been the desire for the program content owners to minimize the cost of distribution, often via satellite. There are a number of different operating points and encoding standards used for this distribution, which will have varying effects on the subsequent emission-encoded picture quality.

The system definition with concatenation is shown below:

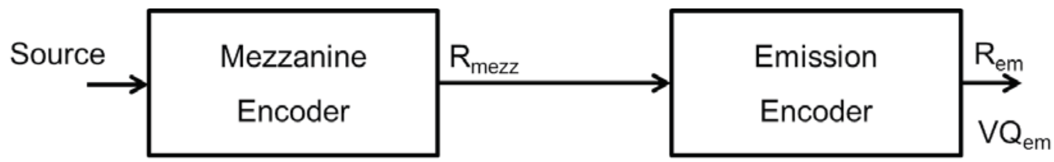


Figure 7: System with concatenation

4.1 Concatenation Performance evaluation

4.1.1 Experiment Configuration

The set-up for evaluation of the effects of concatenation is shown in Figure 8 below. The quality compared to the original source was measured with various configurations of mezzanine link type, with the emission encoder then varied over a range of bit rates to determine the characteristics resulting from each.

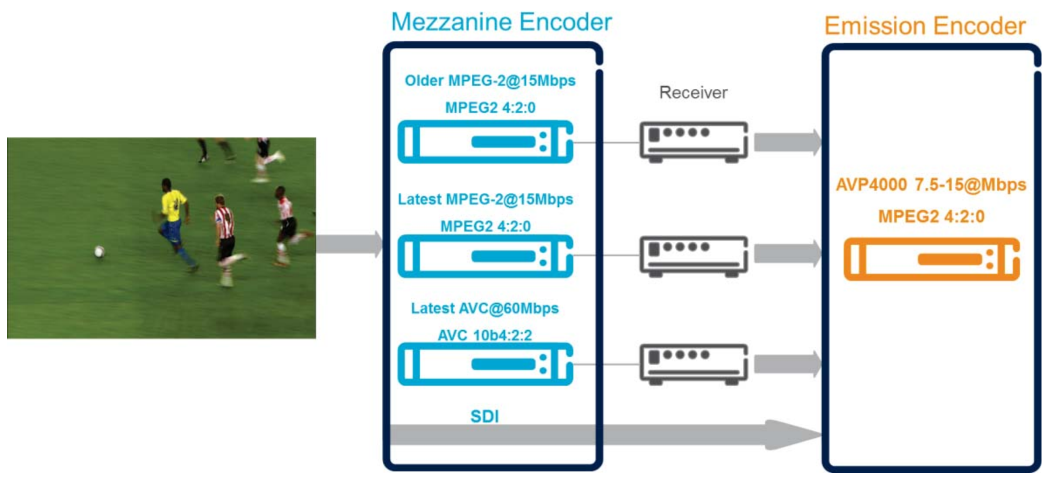


Figure 8: Performance Evaluation Test Configuration

4.1.2 Experiment Content

The experimental results in this document cover a range of types of content, so the content complexity dependency can be evaluated. 4 clips were used, as below:

Table 1: Clip definitions

Clip name	Complexity	Description
Saints	High	Fast-moving sports
Vegas	Easy-moderate	Studio-like content
Rio Carnival	High	High spatial complexity with motion
Gladiator	Moderate	Movie high motion

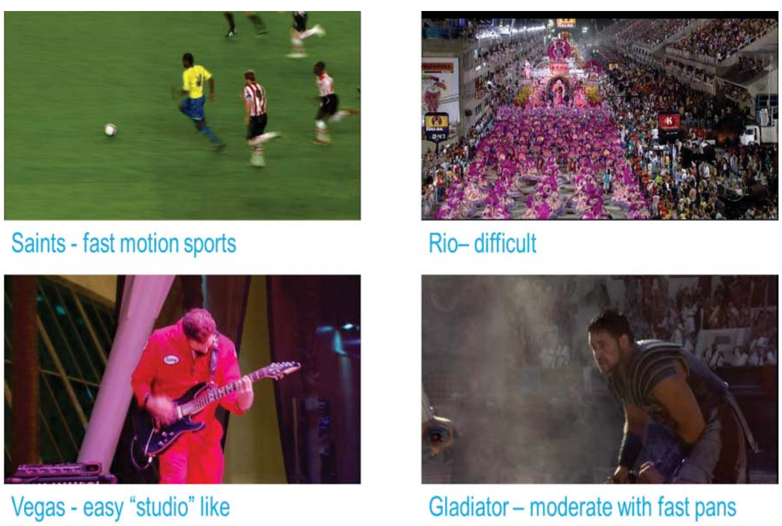


Figure 9: Experimental clip content

4.1.3 Experiment Mezzanine Link Formats

Four different mezzanine link formats were used, as summarized below:

Table 2: Mezzanine encoder format definitions

Format name	Profile/Level	Bit rate	Description
SDI	n/a	n/a	Baseband HD-SDI
Older MPEG-2	MP@HL	15 Mbit/s	Representative of many existing source encoders
Latest MPEG-2	MP@HL	15 Mbit/s	Represents the effect of simply updating the mezzanine encoder but keeping the same format
Latest AVC	422P@HL	60 Mbit/s	Represents a high-quality mezzanine link using the latest MPEG-4 AVC encoder

4.2 Concatenation Performance Results

4.2.1 Effect of Content Type

Figure 10 below shows the effect of concatenated encoding on different types of content. The emission encoder operating point range has been selected to be in the range of interest for the latest developments in QAM utilization.

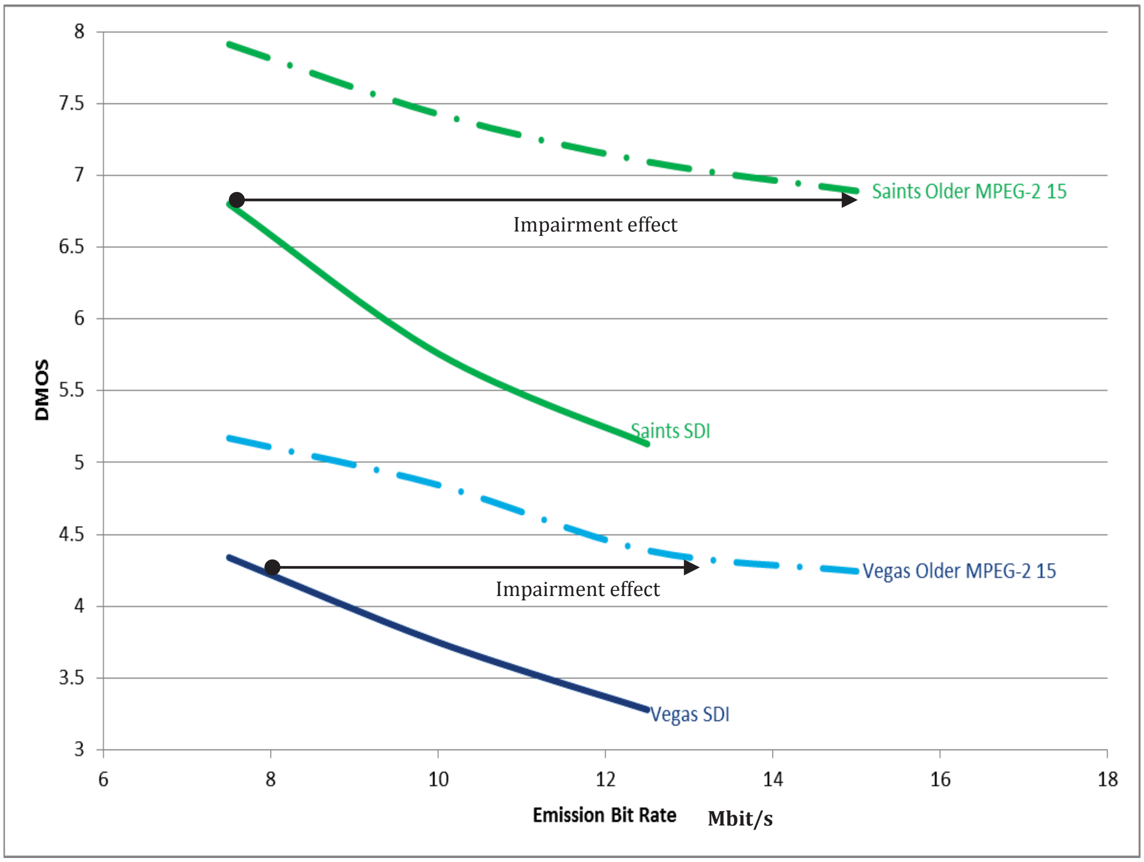


Figure 10: Quality vs bitrate for emission encoder and the impairment effect of a preceding encode stage

Figure 10 compares a commonly existing scheme, where the source is encoded as MPEG-2 at 15 Mbit/s, using an older generation encoder, against the same content but encoded directly from HD-SDI. As can be seen, the degradation effect on the final emission quality is significant for both easy and difficult content; however both the absolute levels of quality and the degree to which it is impaired are dependent on content complexity.

For the Saints (complex) clip, the effect of the preceding encode stage is such that to match the quality of an 8 Mbit/s encoded straight from HD-SDI, the emission encoder would need to be set to around 15 Mbit/s.

For the Vegas (less challenging) clip, the effect of the preceding encode stage is such that to match the quality of an 8 Mbit/s encoded straight from HD-SDI, the emission encoder would need to be set to around 13 Mbit/s.

Note also that the Vegas clip quality is already significantly higher (represented by a lower DMOS score) than the Saints clip, so the concatenation impairment for the Saints clip is both more severe and is applied to a quality that is already at the lower end of the range. Most operators are under pressure to maintain or improve quality, while at the same time reducing bit rate per service.

4.2.2 Effect of Mezzanine Link Compression

The effect of the mezzanine link formats are shown below. For each clip, four different mezzanine formats were used, and the quality versus bit rate for the final emission encoder was established.

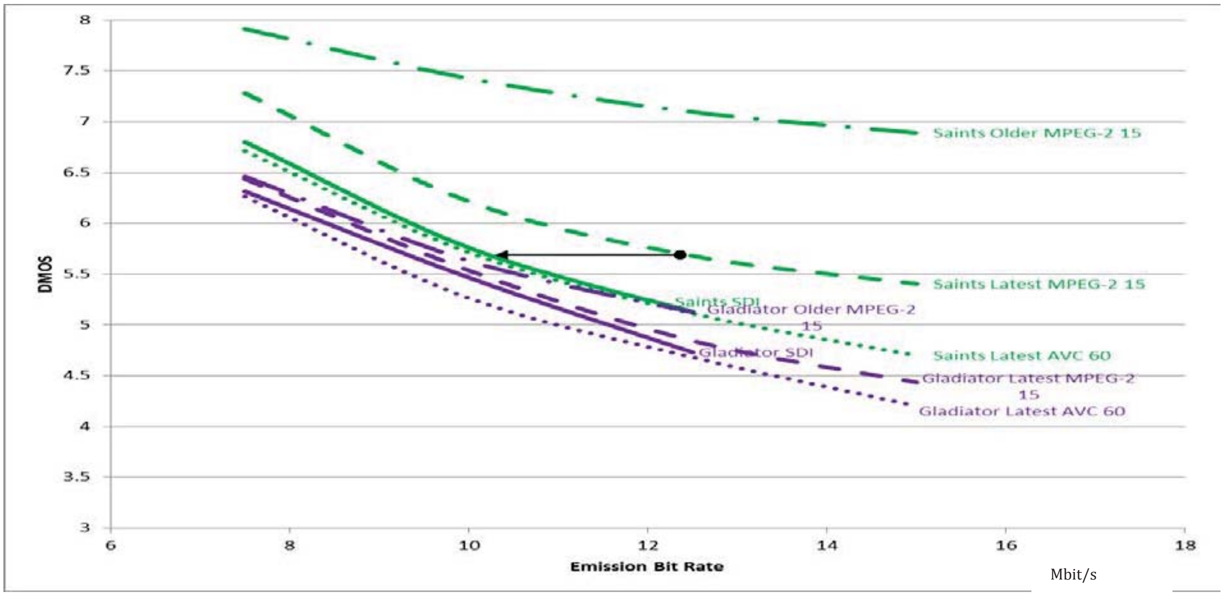


Figure 11: Comparison of mezzanine link effects for Saints and Gladiator clips

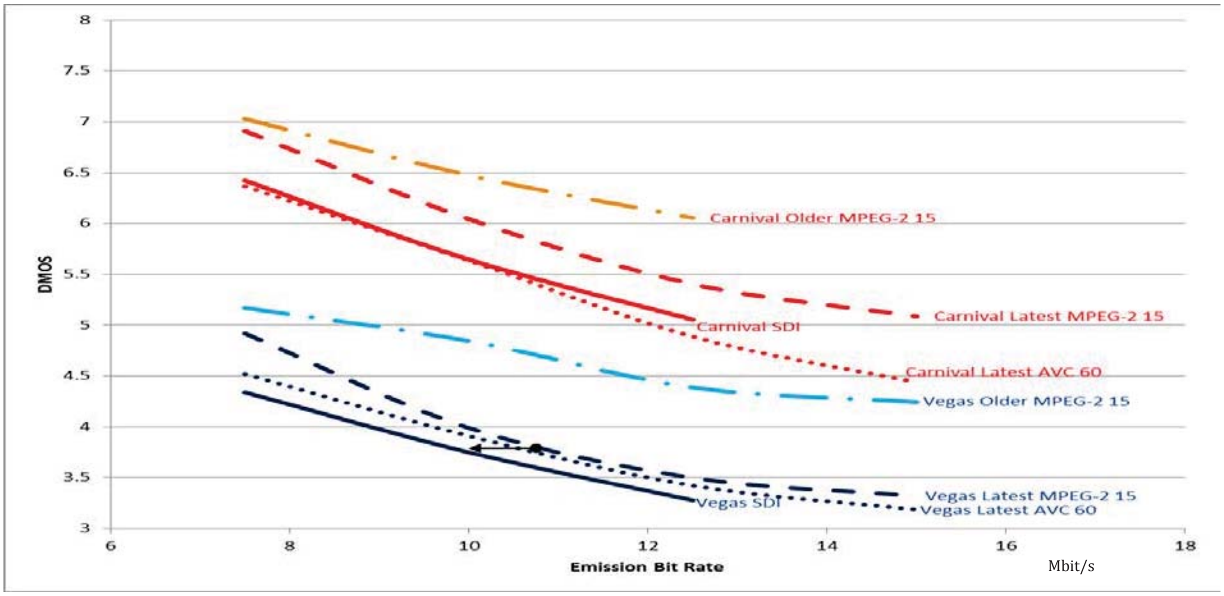


Figure 12: Comparison of mezzanine link effects for Carnival and Vegas clips

Figure 11 and Figure 12 show the effect that the mezzanine link has on the quality (and therefore potential bit rate) of the final emission encode.

In general, many deployed source encoders will lie between the “Older MPEG-2 15” line and the “Latest MPEG-2 15” line.

In all cases, the final emission encoder is a high-performance MPEG-2 HD encoder (Ericsson AVP4000), representing the state of the art available MPEG-2 HD compression technology at the time of writing.

As before, the vertical scale is DMOS, which is an objective measure that aims to model the subjective quality. Lower numbers represent a better observed quality.

4.3 Interpreting the Results

In all cases, the best achievable overall performance is when HD-SDI is used as the source format in conjunction with a high-performance emission encoder. That is the solid line labeled “SDI”. However there are a few cases where the presence of a high bit rate MPEG-4 AVC mezzanine link appears to provide a marginal improvement over HD-SDI. This is most likely due to the application of de-blocking filters and because encoding acts as an entropy reduction process, reducing the detail for the subsequent stage, however the verification of that is outside the scope of this paper.

We can then interpret from the other curves what has been the impact of the mezzanine format encode.

The following initial observations can be made:

- The mezzanine link does have a material impact on the emission quality, unless the operating point of the mezzanine link is much higher than that of the emission encoder
- At MPEG-4 AVC bit rates of around 50-60 Mbit/s, the impairment due to the mezzanine stage becomes negligible, performance is close to encoding directly from the original baseband feed.
- More difficult content is impacted more severely by a constrained bit rate in the mezzanine link, again when measured as an effect at the output of the emission encoders.

4.3.1 Effect on Complex Content

The impact of an older generation MPEG-2 encoder on complex content is such that it limits the available performance regardless of the final (high performance) emission encoder bit rate. This can be seen in Figure 11 with the curve “Saints Older MPEG-2 15”. As can be seen, at the emission encoder bit rate increases, the DMOS score tends to an asymptote that is substantially higher than is achievable even at low emission encoder bit rates when HD-SDI was used.

It is therefore likely that in some existing systems, the performance-limiting stage may well be the mezzanine link encoder rather than the final emission encoder.

Assuming, instead, that the starting point is a latest generation mezzanine encoder running at 15 Mbit/s MPEG-2, we can see that changing to MPEG-4 AVC and running at 60 Mbit/s will result in a reduction of required bit rate from 12.5 Mbit/s to approximately 10 Mbit/s, for the same DMOS score (shown by the double arrow-headed line). This represents a 20% reduction in required bit rate for the emission encoder, and therefore an increase in the number of services in the same bandwidth by a factor of 5/4, i.e. 25%.

Most importantly, it is the most complex content that determines the bit rate that is needed when using a single bit rate CBR for all services in an SDV system and it is that content that benefits most from the provision of a better quality mezzanine feed.

4.3.2 Effect on Easier Content

As we can see from Figure 12, with less challenging content, the effect of the mezzanine link is reduced.

The curves for Vegas show that for the same change (i.e. from a latest generation MPEG-2 encoder at 15 Mbit/s to MPEG-4 AVC at 60 Mbit/s), the reduction is significantly less, approximately 10% in the scenario above.

However, since the absolute video quality is already much higher than for the complex content, the smaller reduction does not affect the potential to operate at a lower bit rate, since reducing the bit rate further still produces a much lower DMOS score than for the complex content.

However, we can also see that an older generation mezzanine encoder ("Vegas Older MPEG-2 15") does materially impair the performance of the overall link, meaning that the efficiency of the emission encoder can still be improved by a better mezzanine encoder, even at the same operating point.

An important consideration, however, is that movie channels can also have advertisements, trailers and other content that are not as low complexity as the movie content itself, so when selecting an operating point for such a channel, it is important to consider the most complex content, rather than the majority, in order to allow emission bit rates to be reduced without unintended consequences when the content temporarily becomes more complex.

4.4 High-Performance Emission Encoders

High performance emission encoders provide the most direct way to improve efficiency, assuming that the mezzanine encoder performance and configuration is not so impaired that it is defining the best achievable quality.

Emission encoder performance has recently significantly increased and the latest high-performance encoders can add further bit rate efficiency. The combination together can allow an emission encoder bit rate reduction to around 9 Mbit/s with equivalent performance.

For a Switched Digital Video (SDV) system where encoders are operated at constant bit rate, the combination of improved mezzanine feeds and high-performance emission encoders can easily translate into the difference between 3 and 4 HD channels per QAM. Since this is CBR, there can be any mix of content complexity within the QAM.

For statistically multiplexed systems, there is additional statistical multiplexing gain efficiency, potentially allowing at 5 HD channels or more depending on content mix within a given QAM.

5 Deployed System Case Study

A real world example of a transition involving both emission and mezzanine encoder changes was tested at Time Warner Cable. The system involved using 60 Mbit/s MPEG-4 AVC as the mezzanine format, as well as a change to the latest state of the art emission encoders for MPEG-2.

The test system was for an SDV based system, so all services operate as CBR MPEG-2 HD. The combination of changes to both parts of the chain allowed an increase in the number of services per QAM from 3 to 4, so overall an improvement of 33% in terms of channels per QAM, while resulting in at least as good video quality as before the change. The resulting emission encoder video bit rate was tested down to 9.5 Mbit/s.

5.1 Other Considerations

5.1.1 10-bit Mezzanine Format

10-bit encoding can provide improved video quality (reduced contouring effects) regardless of bit rate. As a result, wherever possible, 10-bit quantization should be used for the mezzanine links.

Typically, this does not materially affect performance with the most critical content, however when the content is not as challenging, it does offer the potential for improvement in graduated areas [1].

5.1.2 4:2:2

4:2:2 compression provides an increased vertical bandwidth for chrominance information. In particular, repeated 4:2:0 to/from 4:2:2 conversion can caused vertical smearing effects on the chrominance. This is most typically visible for horizontal regions with high saturation, such as logos and overlays, where the chroma may bleed from the overlay up/down into the main content and the overlays can become de-saturated towards the top and bottom of their regions.

As a result, the use of 4:2:2 for mezzanine links is preferable when possible.

5.1.3 JPEG-2000

It is useful to understand the relative merits of using intra-only encoding (i.e. JPEG-2000 or AVC-I) for the mezzanine links. Comparisons are somewhat difficult, because due to the nature of intra-only encoding, there is no dependency on motion, which is a significant factor for compression using prediction between pictures. As a result, the type of content that challenges intra-only compression tends to differ from that which challenges intra- and inter-coded compression.

Since JPEG-2000 artifacts are different in nature from MPEG-4 AVC or MPEG-2 artifacts, a subjective (rather than objective) comparison has been made between MPEG-4 AVC at 60 Mbit/s to identify a subjectively equivalent bit rate for JPEG-2000. The conclusion from this was that approximately 120 Mbit/s was needed for JPEG-2000 for equivalent performance compared to MPEG-4 AVC.

TWC evaluated JPEG-2000 as part of the analysis. Using a subjective review, it was determined that AVC at 60 Mbits/sec was visually as satisfying (if not better) than JPEG-2000 at 120 Mbits/sec.

AVC-I was not evaluated in these tests, however previous analysis has shown a small efficiency benefit compared to JPEG-2000.

5.2 Quality and Video Quality Measurement

The Human Visual System is a complex combination of biological and psychological processes. Ultimately, the only true test of quality is performed as a subjective assessment by a number of viewers [2]. In the interest of practicality, however, there remains a need to approximately measure the subjective quality using an objective measurement that is repeatable and reasonably accurate.

There is no perfect objective measurement that reflects human visual perception; however, there is a range of objective measurements, a few of which are summarized below. In this paper, the MS-SSIM based DMOS measurement has been used, as one of the more representative objective measurements available and does not require proprietary test equipment.

5.2.1 PSNR

Peak Signal-to-Noise Ratio (PSNR) measures amount of noise in test video relative to the reference video. It is a raw measure of distortion that does not take into account the any psycho-visual weighting.

As a result it is not a method for measuring subjective quality.

5.2.2 JND

Just Noticeable Difference (JND) uses spatial and temporal information to measure subjective quality. It has a good correlation with subjective viewing assessment; [3] however it uses a proprietary algorithm and is only available in a limited number of tools.

The JND scale does not represent directly the quality or impairments in measured video but represents ratio of viewers that would be able to differentiate between compressed and original source to those who would not be able to see the difference

5.2.3 DMOS

Differential Mean Opinion Score based on MS-SSIM (Multi-Scale Structural Similarity) uses discrimination of impairment to the structure of the image rather than raw differences. MS-SSIM based DMOS exhibits a good correlation with subjective assessment, however requires appropriate calibration for meaningful use.

One advantage of MS-SSIM metric is that it is available in public domain and has been adopted into many PQ measuring systems. On the other hand the metric originates from image quality rating metrics and does not directly use temporal information and is computed for individual frames so does not offer a temporal element.

6 Conclusions

- High quality mezzanine links have the potential to significantly improve emission encoder efficiency compared to moderately compressed distribution feeds.
- Near-transparency can be achieved with mezzanine links using MPEG-4 AVC at around 60 Mbit/s
- Near-transparency can be achieved with mezzanine links using JPEG-2000 at around 120 Mbit/s
- Combining high performance emission encoding with high quality mezzanine links can significantly improve the efficiency of mapping HD services into QAMs.
- Either MPEG-4 AVC or intra-only formats can successfully be used, however there are no incremental benefits beyond reduced latency for intra-only compression; intra-only compression requires approximately 2x the bit rate of MPEG-4 AVC for comparable performance.

Bibliography

[1] M. Compton, "10-bit High Quality MPEG-4 AVC Video compression," in NAB Broadcast Engineering Conference Proceedings, April 2008.

[2] ITU-R BT.500, "Methodology for the subjective assessment of the quality of television pictures," Jan 2012.

[3] ITU.-T J.144, "Objective perceptual video quality measurement techniques for digital cable," 2004.

Abbreviations & Acronyms

422P@HL	4:2:2 Profile at High Level
AVC	Advance Video Coding (ISO/IEC 14496-10 or ITU-T H.264)
CBR	Constant Bit Rate
DMOS	Differential Mean Opinion Score
HD	High Definition
HD-SDI	High Definition Serial Digital Interface (SMPTE 292M)
JND	Just Noticeable Difference
ML@HL	Main Profile at High Level
MPEG-2	ISO/IEC 13818-2
MS-SSIM	Multi-Scale Structural Similarity Index
PQ	Picture Quality
QAM	Quadrature Amplitude Modulation
Rem	Emission Bit Rate
Rmezz	Mezzanine Link Bit Rate
SDI	Serial Digital Interface (includes HD-SDI)
SDV	Switched Digital Video
VQem	Video Quality at emission