WHITEPAPER

Choosing a Per-Title Encoding Technology

by Jan Ozer (Contributing Editor Streaming Media Magazine)

Introduction

At Streaming Media East 2019, I presented my fourth analysis of per-title encoding technologies which I started reviewing at Streaming Media conferences and writing about for Streaming Media Magazine back in 2017. Bitmovin was the clear winner.

After the show, Bitmovin asked me to write a white paper detailing the analysis. The Streaming Media East Comparison included seven technologies, but Bitmovin wanted the paper to focus on four, their own, capped CRF, Mux, and Elemental.

Otherwise, Bitmovin's sole directions were that the paper be open and impartial, and objective, auditable, and repeatable. Everything else was my work.

You can download the presentation from Streaming Media East here: http://bit.ly/PT_SME_2019. In that analysis, Elemental was Unsub 1, though I reran the files with direction from Elemental for this analysis. While I discuss some aspects of the scoring methodology herein, I explain them fully in this downloadable presentation Scoring Explanation.

My goal (and Bitmovin's directive) was to create a document that would help you understand what per-title technologies do and factors to consider when choosing among them. If you're looking to choose a per-title technology, or even better, to test some systems, you'll get the thoughts I've synthesized while testing over eleven different per-title technologies over four different comparisons.

In the interest of full disclosure, I have consulted with Bitmovin before and tested their system privately in 2017 and 2018 as I have with several companies offering per-title technologies. Note that the analysis changed significantly since that time and that I didn't share these changes with Bitmovin. Still, Bitmovin was familiar with the methodology used herein, though to be fair, all previous analyses have been available for download and review by anyone. So, this analysis shouldn't be a surprise to any per-title developer.

In addition, Bitmovin was the only company where I analyzed a single set of files, delivered six days before the Streaming Media presentation. With Elemental, I analyzed one set of files for the Streaming Media presentation, and then updated the encoding parameters for the files presented herein. With Mux, we tested two sets of files, one that I produced using their public cloud encoder; the second set shown herein which Mux prepared for us.

Summary of Findings

Table 1 shows the results with Bitmovin winning by a substantial margin. Notably, Bitmovin provided the largest improvement to QoE while delivering the second most storage efficiency. Otherwise, Bitmovin was first or second in every category which are explained further below.

	VMAF Accuracy	Storage Efficiency	QoE	Ladder Integrity	Decision Making	Scoring	Overall	Rank
Bitmovin	2	2	1	2	1	1	10	**1**
Capped CRF	3	3	2	3	2	2	17	**2**
Elemental	1	4	3	4	2	3	20	**3**
Mux	4	1	4	1	4	4	25	**4**

Table 1: *Overall rankings place Bitmovin in first place.*

What was particularly impressive was that Bitmovin performed well with all types of videos which you can see in Appendix I. With animations, PowerPoint and Camtasia-based videos, and simple business videos, Bitmovin reduced the data rate while improving quality significantly. With more complicated movie-ish clips, Bitmovin maintained data rate and quality, while for fast moving sports clips, Bitmovin increased the data rate and quality, however slightly.

In the scoring parlance detailed below, Bitmovin made no Bad or Awful decisions, made no errors, and suffered no losses or catastrophes, while accumulating the most home runs. This consistency and error-free performance makes the Bitmovin system very easy to recommend for companies seeking a reliable way to maximize both the QoE and bandwidth of their VOD videos. The only caveat is a minor one-Bitmovin charges 10% extra for per-title encoding.

After a brief discussion of per-title technologies, I jump directly into the analysis. At the end of this paper, I briefly discuss each technology and describe how I created the files analyzed in this paper.

Taxonomy of Per-Title Technologies

Before digging into our test description let's explore the two basic ways that per-title encoding technologies operate, which I'll call "in-rung" and "complete ladder." In-rung technologies start with a fixed encoding ladder and optimize each rung of that ladder individually. So, if you start with seven rungs you finish with seven rungs. In our comparison, Elemental and capped CRF are in-rung technologies.

Complete ladder technologies analyze each video and create a unique ladder for each video, changing both the number of rungs and the resolution of those rungs. Bitmovin and Mux are both complete ladder technologies, though Bitmovin has much more flexibility. That is, depending upon the source, Bitmovin created from three-to seven rungs, with much diversity between rung resolution. In all cases with the second group of files we analyzed, Mux delivered four output files with minor variations in resolution.

While in-rung technologies have their advantages, they typically don't perform well in VOD trials where the ability to customize the complete ladder is more effective. For example, with simple-to-encode clips, a complete ladder technology might create only three rungs, saving both encoding and storage cost. When encoding animations and other synthetic videos, complete rung ladders might deploy larger resolution rungs which produce better quality than lower resolution rungs. In both cases, in-rung technologies produce the same number for rungs at identical resolutions.

Test Description

For a description of the clips and scoring mechanism, please download and refer to the Scoring Explanation document. At a high level, the analysis process goes as follows:

- Encode test clips to the baseline ladder; measure bitrate and compute VMAF, SSIM, and PSNR (Explanation, page 12).
- Encode clips using per-title technology; measure bitrate and compute VMAF, SSIM, and PSNR (Explanation, page 12).
- Assign rungs from per-title technology to fixed title for comparison purposes (Explanation, page 13).
- Compute metric differential and allocate based upon distribution percentage of that rung (Explanation, page 13).

With these results in hand, the analysis begins. The first analysis compares the per-title top rung to the baseline top rung to determine whether the per-title technology made a good decision.

	Decision Making				
	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions
Bitmovin	0	13	5	3	0
Capped CRF	0	7	11	3	0
Elemental	0	7	11	3	0
Mux	0	8	0	0	13

Table 2: *Analyzing the decision making of each technology (Explanation page 16).*

This decision largely depends upon whether the per-title technology decreased VMAF rates in the baseline file from above 93 to below, or whether they boosted rates below 93 in the baseline file higher. In this regard, note the three of Mux's 13 Awful decisions had VMAF scores of 92.5 or higher, with one at 92.92. Under the existing scoring mechanism, these have a disproportionately high cost as compared to a VMAF rating of 92.01. Of course, the best way to avoid this issue is to avoid Awful decisions, which all three other technologies were able to do. In addition, Mux dropped top rung VMAF from above 93 to as low as 82.92, 87.08, and 86.58 and failed to boost the data rates of the three hardest to encode files, which all other technologies did, resulting in the three Great decisions shown.

The statistics shown in Table 3 partially explain Mux's poor decision making and its implications. Integral to each per-title encoding technology is a measure of encoding complexity. The test clips included a range of clips with different complexities from PowerPoint-based tutorials to soccer matches. The other three technologies delivered a standard deviation of under 1.5 VMAF points which means an accurate gauge of complexity across this range. Mux's score of 3.68 indicated that it had issues assigning the effective complexity of the various clips. The graph in Figure 1 below also shows how Mux's gauge of encoding complexity was out of step with all other technologies.

	Overview	
	1080p VMAF Std Deviation	Top Rung Impact
Bitmovin	1.47	0.47
Capped CRF	1.49	0.61
Elemental	1.46	0.04
Mux	3.68	-3.95

Table 3: *VMAF Standard Deviation and top rung impact (Explanation, page 14).*

Looking back at Table 3, the top rung impact is critical because 72.6% of viewers watch that stream. Most producers want a per-title technology that improves hard-to-encode clips and reduces the bitrate of easy-to-encode clips while maintaining a similar quality. Unlike all other technologies, Mux dropped the average VMAF score from 95.44 to 91.49, a level which some videos may begin to show artifacts.

Storage and Streaming and Metrics

Table 4 shows how each per-title technology impacted storage and streaming bitrate as well as overall metric score. In particular, though Bitmovin was fairly thrifty regarding all three storage categories, it provided the highest boost to all three metrics by far. Conversely, though Mux produced the most efficient scores in Storage and Streaming, the overall effect on quality was very severe.

	Storage and Streaming			Metrics		
	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF
Bitmovin	69,767	11,680	3,678	21.38	0.05	24.76
Capped CRF	17,002	-3,069	13,125	14.08	0.03	10.68
Elemental	7,051	-9,122	21,941	7.85	0.00	6.16
Mux	141,085	27,213	0	-20.93	-0.08	-80.48

Table 4: *Storage and Streaming and cumulative metric adjustments (Explanation pages 14-15).*

Table 5 shows the impact of each per-title technique on the encoding ladder and how in-rung and complete-ladder technologies differ. The baseline ladder has seven rungs for each video. Both Bitmovin and Mux are complete-ladder technologies with Mux producing four rungs per video and Bitmovin from three to seven. This means fewer rungs to encode, which may save some encoding costs. Capped CRF and Elemental are both in-rung technologies, so all per-title ladders have the same number of rungs.

Errors relate to the integrity of the encoding ladders produced by each technology. Briefly, Apple recommends that rungs should be between 1.5 – 2x apart to ensure proper operation. Intuitively, if you have too many rungs too close together your encoding costs are unnecessarily high, and you may experience stream switches that don't deliver noticeable quality improvements. If too far, you could strand some viewers at unnecessarily low bitrates degrading their quality of experience.

	Ladder	
	Rungs Saved	Errors
Bitmovin	26	0
Capped CRF	0	30
Elemental	0	41
Mux	63	16

Table 5: *Impact on encoding ladder (Explanation pages 15).*

To identify errors, I counted each rung that was exceeded the 1.5 - 2x recommendation by more than 10%. To be sure, in the vast majority of cases these are unlikely to produce any playback issues. Still, if you were creating the ladder by hand, you would avoid these errors, and clearly Bitmovin, which created a custom ladder for each video, was able to avoid any errors.

Scoring

Scoring is shown in Table 6 (and explained on pages 16-18. As you might suspect, an Awful decision usually presages a Catastrophe and eight of Mux's 13 Awful decisions produced Catastrophes. At the other end of the spectrum, Bitmovin's technology produced eight home runs, two wins, and no losses or catastrophes.

	Scoring				
	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	2	8	0	0	11
Capped CRF	4	4	1	0	12
Elemental	2	2	2	0	15
Mux	2	0	9	8	2

Table 6: *Comparable scoring.*

Synthesis

Table 7 synthesizes the results for each category of the analysis as explained in the Procedures document. Lower is better for each category and overall, and as you can see, Bitmovin ranked first.

	VMAF Accuracy	Storage Efficiency	QoE	Ladder Integrity	Decision Making	Scoring	Overall	Rank
Bitmovin	2	2	1	2	1	1	10	**1**
Capped CRF	3	3	2	3	2	2	17	**2**
Elemental	1	4	3	4	2	3	20	**3**
Mux	4	1	4	1	4	4	25	**4**

Table 7: *Scoring synthesis (page 18).*

Pricing

Beyond the number of rungs in the ladder, there are other noteworthy aspects of pricing which I'll cover technology by technology.

- Bitmovin - Per-title encoding boosts encoding pricing by 10%.
- Capped CRF - Capped CRF is a single-pass technology, so if you're moving from a two-pass technology, you'll save encoding costs by reducing encoding time.
- Elemental - No price premium.
- Mux - No price premium.

Technology Descriptions

The following sections briefly discuss the different technologies from an implementation perspective and detail how I created/obtained the test files.

Bitmovin

Bitmovin is a complete-ladder technology where you upload the file and let the encoder make all the decisions. You can set parameters like minimum and maximum data rate, specific resolutions that must be produced, and others, but we didn't in this case; we just uploaded the files and took what the system delivered us.

Though Bitmovin doesn't have an audience-adaptive component (see Mux review), Bitmovin technical staff can manually tune the system for you if you provide details regarding audience composition and effective bandwidths. We didn't do this in this case; we just used the standard output from their encoder.

Capped CRF

Capped CRF was the only DIY (do it yourself) technology reviewed and it was included because many encoding professionals are familiar with the technique. Like the baseline files, I produced the capped CRF output via FFmpeg using this command string as modified for resolution and bitrate.

ffmpeg -i Tutorial_1080p.mp4 -c:v libx264 -crf 22 -g 60 -keyint_min 60 -sc_threshold 0 -maxrate 9000k -bufsize 9000k -an Tutorial_1080p_CRF.mp4

As noted above, capped CRF is a single-pass technology so is efficient to produce and easily accessible within FFmpeg. For simple tests, like this one, capped CRF's second place performance is impressive. However, though encoding to capped CRF format is simple, most streaming producers need much more extensive packaging and scalable production, which decidedly are not simple to produce in FFmpeg. For most streaming producers, capped CRF simply isn't an option.

AWS Elemental QVBR

I produced the files using the AWS Elemental MediaConvert GUI and AWS Elemental reviewed all settings and procedures. Like CRF, QVBR has multiple settings. After some testing and conferring with AWS Elemental technical staff I used QVBR Level 8 for the 1080p file and level 7 for all others. At Elemental's recommendation, I set the scaler sharpness to 100 and enabled the adaptive dynamic subGOP setting. Otherwise, I encoded all files using the High-Quality Multi-pass technique and leaving all options other than resolution, data rate, and buffer related options at their default.

As an in-rung technology, Elemental has several advantages over complete-ladder technologies, including fast performance and the ability to work with live applications. For VOD files, in-rung technologies lack the flexibility of complete-ladder systems. All other things being equal, for VOD files, properly developed complete-ladder systems should perform better than in-rung technologies, and that's what we saw with Bitmovin in this instance.

Mux

We looked at Mux's technology twice. The first time, I produced the files using the Mux GUI, outputting HLS fragments that we losslessly concatenated into MP4 files to run the quality measurements. Performance of this version was poor; we measured a 1080p VMAF standard deviation of 7.08, with twelve "Awful" decisions leading to 7 Losses and 6 Catastrophes. Overall, Mux scored a 25 in our synthesized scoring system, placing fourth.

Concerned that we had improperly used the system, we contacted Mux. They confirmed that our settings were correct but informed us that they were updating their per-title functionality. So, we waited and tested again using files delivered by Mux from our source files.

Note that Mux is now an audience-adaptive technology that uses device and bandwidth data to create the optimal encoding ladder. The first set of files didn't use this feature; the second set did. Before running the second encode, we provided our distribution assumptions (see explanation page 13) to Mux and gave them the option to encode using this profile or without audience-adaptation. Mux encoded using "global data generated aggregated across the viewing sessions of our platform."

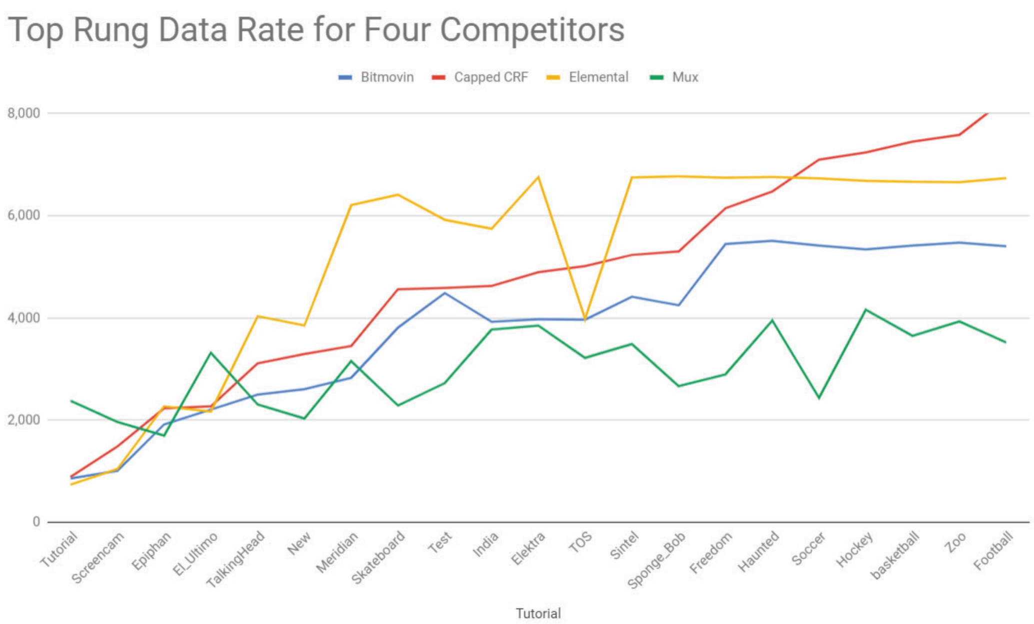


Figure 1: *Top rung data rate for the four per-title technologies.*

As part of the transition to the new system, Mux implemented a fixed four-rung ladder. Though their VMAF standard deviation improved to 3.68, Mux was exceptionally conservative from a data rate perspective as you can see in Figure 8. In fact, Mux was the only technology that didn't increase the data rate of any files, and had several bad misses like in the Skateboard, Soccer, Football, and to a lesser degree, the Basketball clip. This produced 13 Awful decisions.

In addition, the quality of the second rung in Mux's encoding ladder was exceptionally low. Specifically, the average second rung for Mux was a 1112x605 file @ 1.217 Mbps with a VMAF rating of 81. In comparison, the second-rung mux file was 1798x1011@2.2 Mbps for a VMAF rating of 92.83. While Mux's parsimonious approach to rungs saves both encoding costs and storage and bandwidth costs, it does so at a clear cost to QoE.

You can see this in Table 8 which shows the change in VMAF from the baseline files for each rung. Where Bitmovin increased quality on each rung, Mux degraded quality in the top three watched by close to 95% of all viewers and particularly penalized those viewing the second rung, which should still deliver a high-quality experience.

Mux Differential by Rung	VMAF
Rung 1	-3.95
Rung 2	-9.35
Rung 3	-4.05
Rung 4	0.58
Rung 5	-1.53
Rung 6	2.89
Rung 7	29.76

94.55 of viewers watch these three rungs

Bitmovin Differential by Rung	VMAF
Rung 1	0.47
Rung 2	1.71
Rung 3	3.23
Rung 4	3.96
Rung 5	8.67
Rung 6	15.24
Rung 7	30.02

Table 8: *VMAF impact for each rung as compared to Baseline encodes for Mux and Bitmovin.*

As I discuss in Things I'll Do Better Next Time, some of the viewers in the top three rungs are watching on smartphones, where the quality delta is harder to perceive than those watching on TVs or computers. Still, the Mux schema makes those watching on larger, higher quality screens pay a clear price in viewing quality. Not surprisingly, the low quality of the top three rungs produced nine losses and eight catastrophes for the updated Mux schema and another last place finish.

Overall, in speaking to Mux, the priority for their system was fast and inexpensive operation, which the system does deliver. Under this measurement system, however, this comes at a significant cost of QoE, particularly to those viewing the higher rungs of the encoding ladder on larger viewing devices.

Things I'll Do Better Next Time

In our next analysis, I'll try to address these and other issues.

1. VMAF has a phone model that wasn't incorporated into the analysis. Using the phone model, lower resolution files score much higher on smartphones than they do on larger platforms. In the distribution model used for this analysis, smartphones were only 6% of the viewing audience; so not using the phone model had modest effect.

2. The Awful designation, usually triggered by a VMAF score of under 93, needs some fine-tuning, so a score of 92.99 doesn't have a disproportionately significant impact as compared to a score of 93.01. Of course, this only hurts technologies with multiple Awful decisions so the best way to avoid this is to make better decisions.
3. I will explore more sophisticated ways to allocate viewing among the ladder rungs.

Appendix I. Bitmovin Performance by Product Category

Animation	Bitrate	PSNR	SSIM	VMAF	Rungs	Errors	Pass	Good	Bad	Great	Awful	Win	Home Run	Loss	Catas	Draw
El_Ultimo	-1,772	0.984	0.001	1.635	2	0	0	1	0	0	0	0	1	0	0	0
Sintel	-222	0.205	0.002	0.158	1	0	0	1	0	0	0	0	0	0	0	1
Sponge Bob	-328	0.314	0.000	0.674	1	0	0	1	0	0	0	0	0	0	0	1
Average	**-774**	**0.50**	**0.001**	**0.82**	**4**	**0**	**0**	**3**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**2**
Movie-ish	**Bitrate**	**PSNR**	**SSIM**	**VMAF**	**Rungs**	**Errors**	**Pass**	**Good**	**Bad**	**Great**	**Awful**	**Win**	**Home Run**	**Loss**	**Catas**	**Draw**
Elektra	-421	0.191	-0.001	1.020	1	0	0	1	0	0	0	1	0	0	0	0
Freedom	488	-0.095	-0.001	-0.300	1	0	0	0	1	0	0	0	0	0	0	1
Haunted	603	0.986	-0.011	0.513	0	0	0	0	0	1	0	0	0	0	0	1
India	-355	0.289	0.002	0.558	1	0	0	1	0	0	0	0	0	0	0	1
Meridian	-1269	1.022	0.001	1.966	2	0	0	1	0	0	0	0	1	0	0	0
Tears of Steel	-478	0.120	0.000	-0.256	1	0	0	1	0	0	0	0	0	0	0	1
Zoolander	673	-0.022	-0.001	-0.022	1	0	0	0	1	0	0	0	0	0	0	1
Average	**-108**	**0.356**	**-0.002**	**0.497**	**7**	**0**	**0**	**4**	**2**	**1**	**0**	**1**	**1**	**0**	**0**	**5**
Synthetic	**Bitrate**	**PSNR**	**SSIM**	**VMAF**	**Rungs**	**Errors**	**Pass**	**Good**	**Bad**	**Great**	**Awful**	**Win**	**Home Run**	**Loss**	**Catas**	**Draw**
Screencam	-2535	5.787	0.022	5.722	3	0	0	1	0	0	0	0	1	0	0	0
Tutorial	-2948	5.101	0.005	3.309	4	0	0	1	0	0	0	0	1	0	0	0
Average	**-2742**	**5.444**	**0.013**	**4.515**	**7**	**0**	**0**	**2**	**0**	**0**	**0**	**0**	**2**	**0**	**0**	**0**
Other Business	**Bitrate**	**PSNR**	**SSIM**	**VMAF**	**Rungs**	**Errors**	**Pass**	**Good**	**Bad**	**Great**	**Awful**	**Win**	**Home Run**	**Loss**	**Catas**	**Draw**
Epiphan	-1943	2.779	0.002	2.036	2	0	0	1	0	0	0	0	1	0	0	0
New	-1159	0.771	0.003	0.974	2	0	0	1	0	0	0	1	0	0	0	0
Talking head	-1496	0.599	0.000	1.799	1	0	0	1	0	0	0	0	1	0	0	0
Test	-68	0.664	0.002	0.996	1	0	0	0	1	0	0	0	0	0	0	1
Average	**-1167**	**1.203**	**0.002**	**1.451**	**6**	**0**	**0**	**3**	**1**	**0**	**0**	**1**	**2**	**0**	**0**	**1**
Sports	**Bitrate**	**PSNR**	**SSIM**	**VMAF**	**Rungs**	**Errors**	**Pass**	**Good**	**Bad**	**Great**	**Awful**	**Win**	**Home Run**	**Loss**	**Catas**	**Draw**
Basketball	506	-0.189	-0.001	-0.689	0	0	0	0	1	0	0	0	0	0	0	1
Football	581	0.846	0.023	3.034	0	0	0	0	0	1	0	0	1	0	0	0
Hockey	557	0.873	0.005	2.002	1	0	0	0	0	1	0	0	1	0	0	0
Skateboard	-581	0.147	-0.001	-0.214	1	0	0	1	0	0	0	0	0	0	0	1
Soccer	488	0.008	0.000	-0.152	0	0	0	0	1	0	0	0	0	0	0	1
Average	**310**	**0.34**	**0.01**	**0.80**	**2**	**0**	**0**	**1**	**2**	**2**	**0**	**0**	**2**	**0**	**0**	**3**
	-11680	**21.38**	**0.05**	**24.76**	**26**	**0**	**0**	**13**	**5**	**3**	**0**	**2**	**8**	**0**	**0**	**11**

Appendix

www.streaminglearningcenter.com
jozer@mindspring.com
@janozer

How I Tested

About 50 minutes of video in total:

Title	Genre
https://s3.amazonaws.com/pertitle/Basketball 1080p.mp4	Basketball
https://s3.amazonaws.com/pertitle/El Ultimo 1080p.mp4	Simple animated movie
https://s3.amazonaws.com/pertitle/Elektra 1080p.mp4	Movie
https://s3.amazonaws.com/pertitle/Epiphan_1080p.mp4	Screencam and video
https://s3.amazonaws.com/pertitle/Football 1080p.mp4	Harmonic football clip
https://s3.amazonaws.com/pertitle/Freedom 1080p.mp4	Music video
https://s3.amazonaws.com/pertitle/Haunted 1080p.mp4	Movie like video
https://s3.amazonaws.com/pertitle/Hockey 1080p.mp4	Animated movie
https://s3.amazonaws.com/pertitle/India 1080p.mp4	Videos from India
https://s3.amazonaws.com/pertitle/Meridian 1080p.mp4	Meridian
https://s3.amazonaws.com/pertitle/New 1080p.mp4	Test clip
https://s3.amazonaws.com/pertitle/Screencam 1080p.mp4	Screencam only
https://s3.amazonaws.com/pertitle/Sintel 1080p.mp4	Animated movie
https://s3.amazonaws.com/pertitle/Skateboard 1080p.mp4	Skateboard
https://s3.amazonaws.com/pertitle/Soccer 1080p.mp4	Soccer match
https://s3.amazonaws.com/pertitle/Sponge Bob 1080p.mp4	Animated movie
https://s3.amazonaws.com/pertitle/TOS_1080p.mp4	Movie with computer generated content
https://s3.amazonaws.com/pertitle/TalkingHead 1080p.mp4	Simple talking head
https://s3.amazonaws.com/pertitle/Test 1080p.mp4	Mixed talking head and ballet
https://s3.amazonaws.com/pertitle/Tutorial 1080p.mp4	Mixed PowerPoint and video
https://s3.amazonaws.com/pertitle/Zoo_1080p.mp4	Movie footage

Animation
El_Ultimo
Sintel
Sponge Bob
Average
Movie-ish
Elektra
Freedom
Haunted
India
Meridian
Tears of Steel
Zoolander
Average
Synthetic
Screencam
Tutorial
Average
Other Business
Epiphan
New
Talking head
Test
Average
Sports
Basketball
Football
Hockey
Skateboard
Soccer
Average

This encoding ladder as baseline (with FFmpeg). Per-title:

- 2 second GOP, 2 second VBV
- High profile
- 150% upwards
- Unlimited downwards

Width	Height	Profile	Preset	GOP	Data Rate	Max Rate	VBV Buffer	FPS	Audio
1920	1080	High	Medium	2 seconds	4500	9000	9000	Native	None
1280	720				2700	5400	5400		
960	540				1900	3800	3800		
854	480				1350	2700	2700		
640	360				900	1800	1800		
480	270				500	1000	1000		
320	180				250	500	500		

Scoring

	Width	Height	Bitrate	PSNR	SSIM	VMAF
Meridian_1080p_CVBR.m	1920	1080	4,560	43.95	0.964	95.92
Meridian_720p_CVBR.mp	1280	720	2,734	40.02	0.957	90.53
Meridian_540p_CVBR.mp	960	540	1,921	37.99	0.949	85.14
Meridian_480p_CVBR.mp	854	480	1,364	37.07	0.943	81.90
Meridian_360p_CVBR.mp	640	360	905	35.35	0.930	72.73
Meridian_270p_CVBR.mp	480	270	496	33.88	0.913	58.30
Meridian_180p_CVBR.mp	320	180	239	31.36	0.883	24.86
			12,218	**37.09**	**0.934**	**72.77**

Scoring: *Starting point is constrained VBR ladder*

	Width	Height	Data Rate		PSRN	SSIM	VMAF
Meridian_PT_1080p_2839158.mp4	1920	1080	2830	1.95	43.33	0.960	95.17
Meridian_PT_1080p_1545226.mp4	1920	1080	1451	1.87	41.74	0.950	92.09
Meridian_PT_900p_840998.mp4	1600	900	776.8	1.89	39.21	0.939	86.10
Meridian_PT_576p_442631.mp4	1024	576	410.7	1.95	36.75	0.928	77.63
Meridian_PT_432p_240000.mp4	768	432	210.8		34.74	0.910	65.35
			5,679		**39.15**	**0.937**	**83.27**

Scoring: *Get the per-title encode*

Slot the files into the new ladder based upon the rung the viewer would see at each bandwidth @ 110%. Assumed all viewers not in lower rungs could view highest rung

- Never exceeded 6500 kbps or so
- Probably not 100% correct, but only workable assumption

	Width	Height	Data Rate		PSRN	SSIM	VMAF
Meridian_PT_1080p_2839	1920	1080	2830	1.95	43.33	0.960	95.17
Meridian_PT_1080p_1545	1920	1080	1451	1.87	41.74	0.950	92.09
Meridian_PT_900p_84099	1600	900	776.8	1.89	39.21	0.939	86.10
Meridian_PT_576p_44263	1024	576	410.7	1.95	36.75	0.928	77.63
Meridian_PT_432p_24000	768	432	210.8		34.74	0.910	65.35
			5,679		**39.15**	**0.937**	**83.27**

	Width	Height	Data Rate	PSRN	SSIM	VMAF
Meridian_PT_	1920	1080	2,830	43.33	0.960	95.17
Meridian_PT_	1920	1080	2,830	43.33	0.960	95.17
Meridian_PT_	1920	1080	1,451	41.74	0.950	92.09
Meridian_PT_	1920	1080	1,451	41.74	0.950	92.09
Meridian_PT_	1600	900	777	39.21	0.939	86.10
Meridian_PT_	1024	576	411	36.75	0.928	77.63
Meridian_PT_	768	432	211	34.74	0.910	65.35
			9,960	**40.12**	**0.942**	**86.23**

Compare Per-Title Ladder to Original and Allocate

On a rung by rung basis, compute the difference in bitrate and metrics score. Allocate change based upon assumed viewing percentage of each rung (Brightcove white paper ladder 2).

Device type	Usage [%]	Average bandwidth [Mbps]
PC	63.49	14.720
Mobile	6.186	10.609
Tablet	9.165	12.055
TV	21.15	24.986
All devices	100	16.393

		PC	Phone	Tablet	TV	Aggregate
1920	1080	74.16%	61.93%	61.06%	89.25%	71.60%
1280	720	11.33%	20.39%	14.76%	7.42%	13.48%
960	540	7.49%	11.30%	16.41%	2.56%	9.44%
854	480	3.70%	3.34%	5.45%	0.54%	3.26%
640	360	1.70%	1.47%	1.52%	0.17%	1.21%
480	270	0.62%	1.16%	0.52%	0.04%	0.58%
320	180	0.35%	0.26%	0.15%	0.01%	0.19%

Per-title results minus baseline times allocation

Allocation	Data Rate	PSNR	SSIM	VMAF
71.60%	-1,239	-0.444	-0.003	-0.541
13.48%	13	0.446	0.000	0.625
9.44%	-44	0.354	0.000	0.655
3.26%	3	0.152	0.000	0.332
1.21%	-2	0.047	0.000	0.162
0.58%	0	0.017	0.000	0.113
0.19%	0	0.007	0.000	0.078
		0.578	**-0.002**	**1.424**
	-1,269	**1.022**	**0.001**	**1.966**

Scoring

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	6	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

- Greeb is the best, yellow is the worst
- Mux delivered the most file savings and most efficient production
- Mux quality was the lowest of the three (see last two slides)

1080p VMAF Standard Deviation

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	8	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

1080p VMAF standard deviation

- Measures accuracy of quality metric used by per-title technique relating to VMAF
- Lower numbers are better

Top rung impact

- Effecton VMAF score for the top rung

Storage Saved

Storage bandwidth saved over test videos (~50 minutes). Per-title bitrate ladders vs. constrained VBR baseline.

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	8	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

Streaming Bandwidth Saved

Streaming bandwidth saved over test videos (~50 minutes). Higher numbers better.

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	8	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

Wasted Bandwidth

Bandwidth increases in first rung when already over 93 VMAF. So,if VMAF was 94 in baseline file, and 95 in the per-title, the increased bandwidth would be included. Lower numbers better.

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	8	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

Impact on PSNR, SSIM, and VMAF

Overall impact on QoE as measured by PSNR, SSIM, and VMAF

- Higher scores are better with all three metrics
- Negative numbers indicate quality degradation

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidt	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	6	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

Rungs Saved

Rungs eliminated: Started with 7 for each video; Can save encoding and storage costs. Higher numbers better.

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	6	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

Errors

Errors – ladder integrity issues:

- Rungs should be between 1.5 –2x apart to ensure proper operation
- Anytime encoder exceeded this by 10% it was an error
- In most cases, even substantial variations won't cause a playback problem, particularly in the highest or lowest rungs (language is baseball terminology, not descriptive)
- Lower numbers better

	Width	Height	Data Rate		PSRN	SSIM	VMAF
Haunted_1080p_YT.mp4	1920	1080	4,365	2.18	41.22	0.968	89.31
Haunted_720p_YT.mp4	1280	720	2,006	1.64	39.25	0.963	83.70
Haunted_480p_YT.mp4	854	480	1,225	2.67	36.16	0.949	66.93
Haunted_360p_YT.mp4	640	360	460	1.38	34.63	0.941	59.76
Haunted_240p_YT.mp4	426	240	332		31.65	0.924	37.24
			8,387		**36.58**	**0.949**	**67.39**

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	6	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

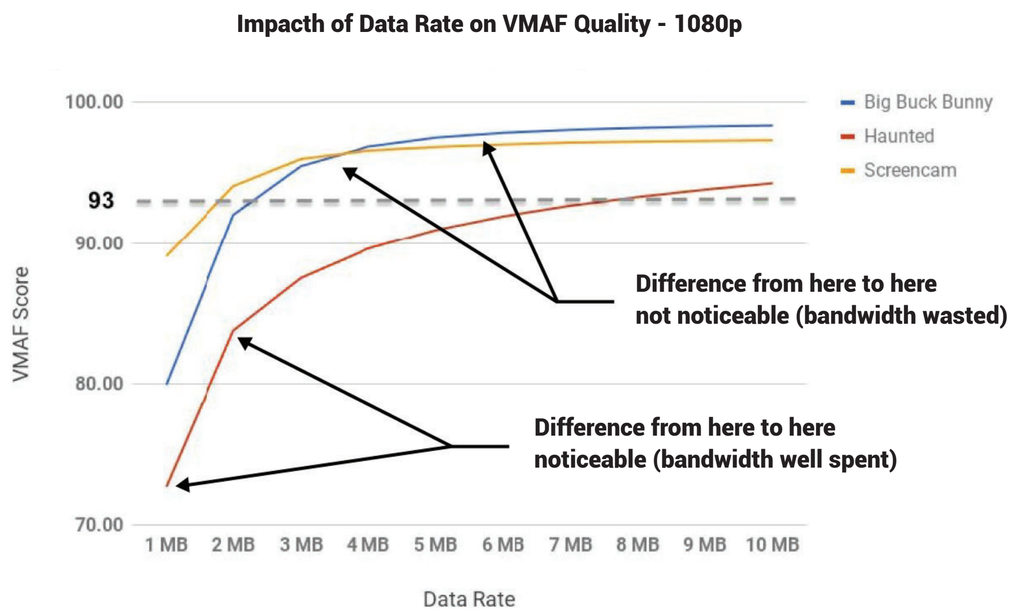
Decision Making

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	6	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

- Pass, Good, Bad, Great, Awful (1080p rung only): Theory: if you were encoding manually, and moving from the original bitrate ladder to per-title, would you increase or decrease the 1080p bitrate?
- **Pass** – if VMAF between 93-95, data rate stays within 95%-105% (in essence, stayed pat)
- **Good decision** – noted, but doesn't directly impact scoring: Start at over 95 VMAF and decrease data rate, but not below 93 (viewer wouldn't notice).
- **Bad decision** – noted, but doesn't directly impact scoring: Increasing the data rate when already over 93 (viewer wouldn't notice).
- **Great decision** – impacts scoring (more later): Increasing data rate when under 93 VMAF
- **Awful decision** – impact scoring (more later): Decreasing the data rate when under 93 VMAF / Decreasing the data rate from above to below 93 VMAF

Working With VMAF

Range 0 – 100. Top rung target – typically 93: 93 delivers video "either indistinguishable from original or with noticeable but not annoying distortion." (bit.ly/VMAF_93); 6 VMAF points = Just noticeable difference.



Working With the Top Rung

	Width	Height	Bitrate	PSNR	SSIM	VMAF			Width	Height	Data Rate		PSRN	SSIM	VMAF
Epiphan_1080p_CVBR.mp4	1920	1080	4,493	51.05	0.989	96.63	**4500**	Epiphan_1080p_YT.mp4	1920	1080	1,449	1.69	47.68	0.984	94.65
Epiphan_720p_CVBR.mp4	1280	720	2,671	41.8	0.984	92.81	**2700**	Epiphan_720p_YT.mp4	1280	720	856	2.04	40.61	0.979	90.41
Epiphan_540p_CVBR.mp4	960	540	1,872	38.97	0.974	87.61	**1900**	Epiphan_480p_YT.mp4	854	480	419	1.83	35.26	0.956	73.30
Epiphan_480p_CVBR.mp4	854	480	1,329	37.8	0.968	84.55	**1350**	Epiphan_360p_YT.mp4	640	360	229	1.84	33.82	0.943	65.30
Epiphan_360p_CVBR.mp4	640	360	892	35.26	0.952	73.59	**900**	Epiphan_240p_YT.mp4	426	240	125		30.51	0.912	35.59
Epiphan_270p_CVBR.mp4	480	270	496	32.79	0.931	56.26	**500**								
Epiphan_180p_CVBR.mp4	320	180	245	30	0.904	25.32	**250**								
			11,998	**38.24**	**0.957**	**73.83**					**3,077**		**37.58**	**0.955**	**71.85**

Essentially, any score over 93 adds little perceptible value:

- Shouldn't penalize drop from 96.63 to 94.65 (viewer wouldn't notice)
- Or reward increase from 96 to 98 (again, viewer wouldn't notice)

So, exclude from scoring calculations:

- Increases in data rate when score already above 93 (bad decision, so no benefit)
- Decrease data rate when both scores above 93 (good decision, no QoE drop)

Include top rung in overall scoring when:

- Increase data rate when below 93 (reward –great decision)
- Decrease data rate from above 93 to below 93 (penalize –awful decision)
- Decrease data rate when below 93 (ditto)

Bitrate	Rez	VMAF	Allocation	Bitrate	PSNR	SSIM	VMAF
10.42%	0	0.14	71.60%	331	0.064	0.001	0.098
-31.00%	0	-1.60	13.48%	-112	-0.061	0.000	-0.216
-2.53%	180	2.58	9.44%	-5	0.063	0.000	0.243
-19.19%	60	-0.49	3.26%	-8	-0.001	0.000	-0.016
-1.67%	120	5.50	1.21%	0	0.008	0.000	0.067
-29.32%	0	-3.27	0.58%	-1	-0.002	0.000	-0.019
-18.77%	0	-1.41	0.19%	0	0.000	0.000	-0.003
					0.072	**0.000**	**0.154**
No top				**205**	**0.007**	**-0.001**	**0.056**

Includes top rung

Doesn't include

When red in spreadsheet, top line counted (great or awful decisions)

Wins, Home Runs, Losses, Catastrophes and Draws

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	8	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

Data Rate	Overall VMAF	Result
Down > 1 Mbps	Down > 1	Loss
	Between -1 - 0	Draw
	Up	Win
-1 Mbps / + 1 Mbps	Down > 1	Loss
	Between -1 - +1	Draw
	Up > 1	Win
Up > 1 Mbps	Down >	Loss
	Between 0 - +1	Draw
	Up > 1	Win
	Up over 1.5	Home Run
	Down over > 3.5	Catastrophe

Initial VMAF	Bitrate	Bitrate	Rez	VMAF	Allocation	Bitrate	PSNR	SSIM	VMAF
98.72	-914	-20.03%	0	-4	71.60%	-654	-1.425	-0.01	-3.142
		-52.45%	-162	-14.966	13.48%	-193	-0.407	-0.004	-2.017
		-32.36%	18	-10.700	9.44%	-59	-0.177	-0.002	-1.010
		-4.42%	78	-2.963	3.26%	-2	-0.014	0.000	-0.096
		-22.98%	8	-7.113	1.21%	-3	-0.012	0.000	-0.086
		-20.68%	-16	-6.059	0.58%	-1	-0.005	0.000	-0.035
		62.38%	74	17.552	0.19%	0	0.003	0.000	0.034
							-2.037	**-0.01**	**-6.353**
		No top				**-911**	**-0.612**	**-0.006**	**-3.211**
						0	0	0	

Based on this number from spreadsheet (seebelow)

Scoring Synthesis

	Overview		Storage and Streaming			Metrics			Ladder		Decision Making					Scoring				
	1080p VMAF Std Deviation	Top Rung Impact	Storage Saved	Streaming Bitrate Saved	Wasted Bandwidth	PSNR	SSIM	VMAF	Rungs Saved	Errors	Pass	Good Decisions	Bad Decisions	Great Decisions	Awful Decisions	Wins	Home Run	Losses	Cata-strophes	Draws
Bitmovin	1.47	0.47	69,767	11,680	3,678	21.38	0.05	24.76	26	0	0	13	5	3	0	2	8	0	0	11
Capped CRF	1.49	0.61	17,002	-3,069	13,125	14.08	0.03	10.68	0	30	0	7	11	3	0	4	4	1	0	12
Elemental	1.46	0.04	7,051	-9,122	21,941	7.85	0.00	6.16	0	41	0	7	11	3	0	2	2	2	0	15
Mux	3.68	-3.95	141,085	27,213	0	-20.93	-0.08	-80.48	63	16	0	8	0	0	13	2	0	9	8	2

- VMAF accuracy: Ranking, lower is better
- Storage efficiency: 1x times storage + 10x times streaming, higher is better
- QoE: 1x times VMAF + 2.5x times PSRN +100 times SSIM, higher is better
- Ladder integrity: Rungs saved minus errors, higher is better
- Decision making: 1x times good decision + 2x times great decision minus 1x time bad decisions and 2x times awful decisions, higher is better
- Scoring: 1x times win + 2x times home run minus 1x time loss minus 2x times awful decisions, higher is better
- Final Adjustments: QoE x 3 because most important metric; Ladder integrity/decision making by .25% because least important
- Pricing mentioned in text

	VMAF Accuracy	Storage Efficiency	QoE	Ladder Integrity	Decision Making	Scoring	Overall	Rank
Comp A	2	2	1	2	1	1	10	**1**
Capped CRF	3	3	2	3	2	2	17	**2**
Comp B	1	4	3	4	2	3	20	**3**
Mux	4	1	4	1	4	4	25	**4**