HEVC: RATING THE CONTENDERS

Jan Ozer
www.streaminglearningcenter.com
jozer@mindspring.com/
276-235-8542
@janozer

Agenda

- Who are the competitors?
- Setting the ground rules
- Results
- Conclusions

Who Was Invited

- Companies asked to participate
  - Intel – distributing through OEMs only
  - Elemental – declined to participate
  - Ittiam – declined to participate – no bandwidth
  - Beamr/Vanguard – declined – bad timing
  - Vantrix – f.265 – no real effort behind technology
  - NTT – declined, primarily sell custom products to specific OEMs
  - MainConcept – yes!
  - X265 – publicly available – yes!

What I tested

- Codecs
  - x264 as baseline
  - Main Concept HEVC
  - x265
  - VP9 (Google)
  - Bitmovin AV1 (more later)
- Focus
  - VOD
  - Encoding time specific
  - 720p, 1080p, 4K
    - Five profiles each
  - Four video files
    - Netflix Meridian
    - Blender TOS
    - Blender Sintel
    - New test clip

Basic Encoding Parameters

720p	Codec	Width	Height	Data Rate	Bits/pixel (30 fps)	Max DR	VBV	Profile	Key-Frame	B-frame	Reference
500 kbps	**HEVC**	1280	720	500	0.0181	550	500	Main	2	3	6
1 mbps	**HEVC**	1280	720	1000	0.0362	1100	1000	Main	2	3	6
2 mbps	**HEVC**	1280	720	2000	0.0723	2200	2000	Main	2	3	6
3 mbps	**HEVC**	1280	720	3000	0.1085	3300	3000	Main	2	3	6
4 mbps	**HEVC**	1280	720	4000	0.1447	4400	4000	Main	2	3	6
1080p											
1 mbps	**HEVC**	1920	1080	1000	0.0161	1100	1000	Main	2	3	6
2 mbps	**HEVC**	1920	1080	2000	0.0322	2200	2000	Main	2	3	6
4 mbps	**HEVC**	1920	1080	4000	0.0643	4400	4000	Main	2	3	6
6 mbps	**HEVC**	1920	1080	6000	0.0965	6600	6000	Main	2	3	6
8 mbps	**HEVC**	1920	1080	8000	0.1286	8800	8000	Main	2	3	6
4K											
4 mbps	**HEVC**	3840	2160	4000	0.0161	4400	4000	Main	2	3	6
8 mbps	**HEVC**	3840	2160	8000	0.0322	8800	8000	Main	2	3	6
12 mbps	**HEVC**	3840	2160	16000	0.0643	17600	16000	Main	2	3	6
16 mbps	**HEVC**	3840	2160	24000	0.0965	26400	24000	Main	2	3	6
20 mbps	**HEVC**	3840	2160	32000	0.1286	35200	32000	Main	2	3	6

Data Rate to +/- 5%

	Sintel			Tears of Steel				New			Meridien	
	120	Actual	%	180	Actual	%	96	Actual	%	180	Actual	%
720p												
500	7,500	7,501	100.01%	11,250	11,341	100.81%	6,000	6,197	103.28%	11,250	10,000	88.89%
1000	15,000	14,677	97.85%	22,500	22,660	100.71%	12,000	12,024	100.20%	22,500	22,510	100.04%
2000	30,000	30,077	100.26%	45,000	45,625	101.39%	24,000	24,255	101.06%	45,000	45,016	100.04%
3000	45,000	45,243	100.54%	67,500	68,326	101.22%	36,000	36,264	100.73%	67,500	67,575	100.11%
4000	60,000	60,230	100.38%	90,000	94,279	104.75%	48,000	48,240	100.50%	90,000	90,153	100.17%

Encoding Details – x265

- Encoded in x265 (not FFmpeg)
- 8-bit version (Main)
- Simple command string created by x265 MulticoreWare
- PSNR/SSIM tuning enabled for all tests

x265_8bit.exe --input TOSN_720p.yuv --input-res 1280x720 --fps 24 --keyint 48 --min-keyint 48 --bframes 3 --ref 6 --bitrate 2000 --vbv-maxrate 2200 --vbv-bufsize 2000 --preset medium --output TOSN_720p_2M.hevc --no-scenecut --tune psnr --psnr --ssim --pass 1 --slow-firstpass

x265_8bit.exe --input TOSN_720p.YUV --input-res 1280x720 --fps 24 --keyint 48 --min-keyint 48 --bframes 3 --ref 6 --bitrate 2000 --vbv-maxrate 2200 --vbv-bufsize 2000 --preset medium --output TOSN_720p_2M.hevc --no-scenecut --tune psnr --psnr --ssim --pass 2

A Note About Tuning

- Tuning doesn't "tune" the codec to produce better PSNR/SSIM scores.
- Rather, "They simply turn off quality enhancement algorithms that are known to degrade objective measurements, although we also know that they improve subjective visual quality. x265 has never been, and never will be "tuned" for optimal psnr or ssim scores."
  - Tom Vaughan, VP and GM, MulticoreWare (developer of x265)
- So, tuned for PSNR testing with x264/x265

Encoding Details – MainConcept

- Encoded with encoder SDK v.1
- Command string and ini files supplied by MainConcept
- Used SABET for encoding
  - Smart Adaptive Bitrate Encoding Technology.
  - Encode source into up to 12 HEVC videos with different resolution and bitrate in one encoding step
  - Primary benefit is encoding time saving with a very slight cost in quality
    - According to MC, my tests did not confirm

Encoding Details – x264

- Encoded in FFmpeg
- Simple command string
- PSNR tuning enabled

```
ffmpeg -y -i TOSN_720p.mp4 -c:v libx264 -preset veryslow -g 60 -keyint_min 60 -
sc_threshold 0  -bf 3 -refs 6 -b_strategy 2 –tune PSNR -b:v 1000k -pass 1  -f mp4 NUL && \

ffmpeg -i TOSN_720p.mp4 -c:v libx264 -preset veryslow -g 60 -keyint_min 60 -sc_threshold
0  -bf 3 -refs 6  -b_strategy 2 –tune PSNR  -b:v 1000k -maxrate 1100k -bufsize 1000k -pass
2  TOSN_720p_1M.mp4
```

Encoding Details – VP9

- Encoded in same version of FFmpeg
- Reviewed by Google for my book, Video Encoding by the Numbers

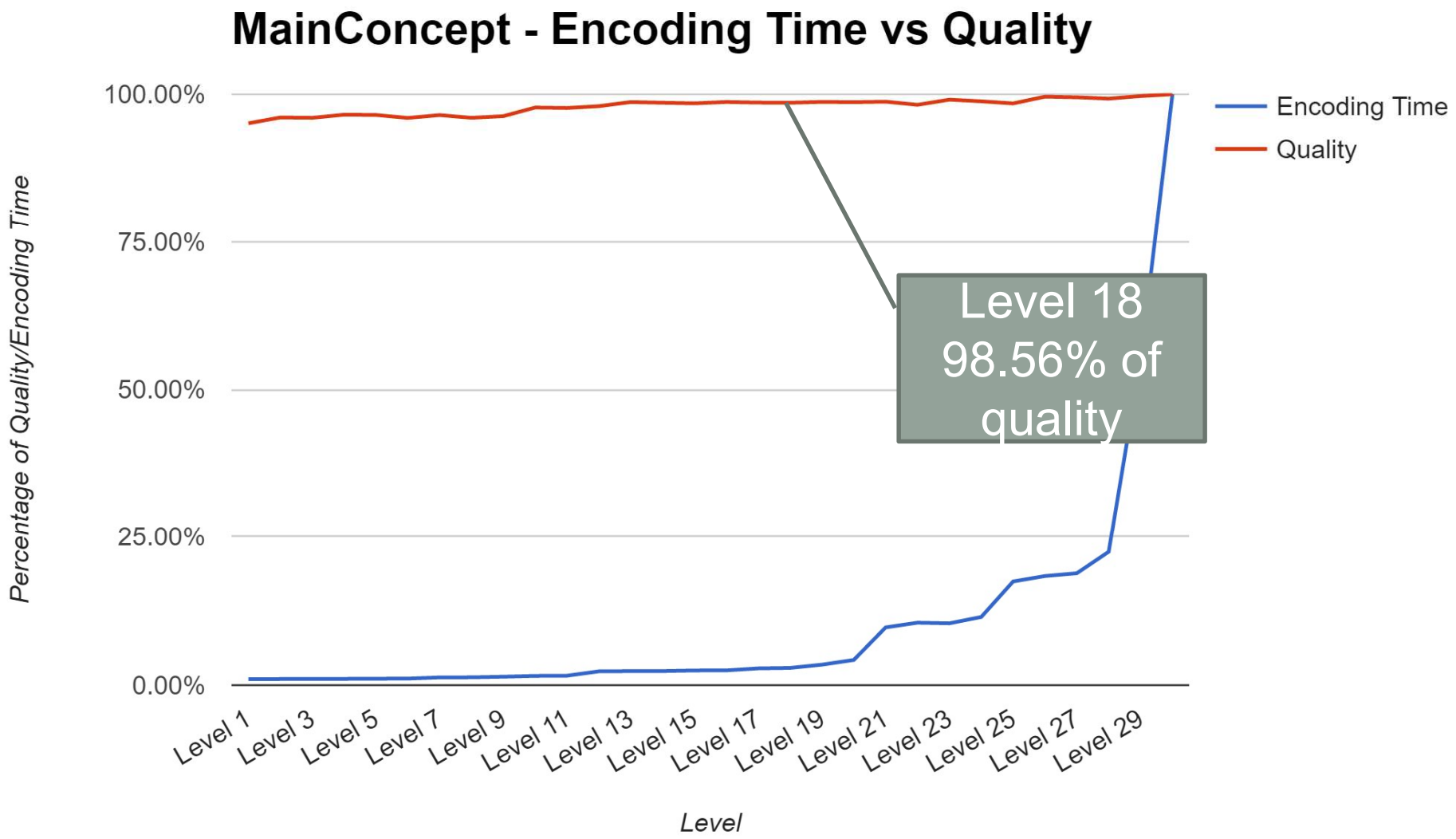
```
ffmpeg -y -i TOSN_720p.mp4 -c:v libvpx-vp9 -pass 1 -b:v 8000K -keyint_min 48 -g 48  -
threads 8 -speed 1 -tile-columns 4 -frame-parallel 1  -f webm NUL && \

ffmpeg -i TOSN_720p.mp4 -c:v libvpx-vp9 -pass 2 -b:v 4000K -keyint_min 48 -g 48  -
threads 8 -speed 1 -tile-columns 4  -frame-parallel 1  -an -f webm sintel_720p_4M.webm
```

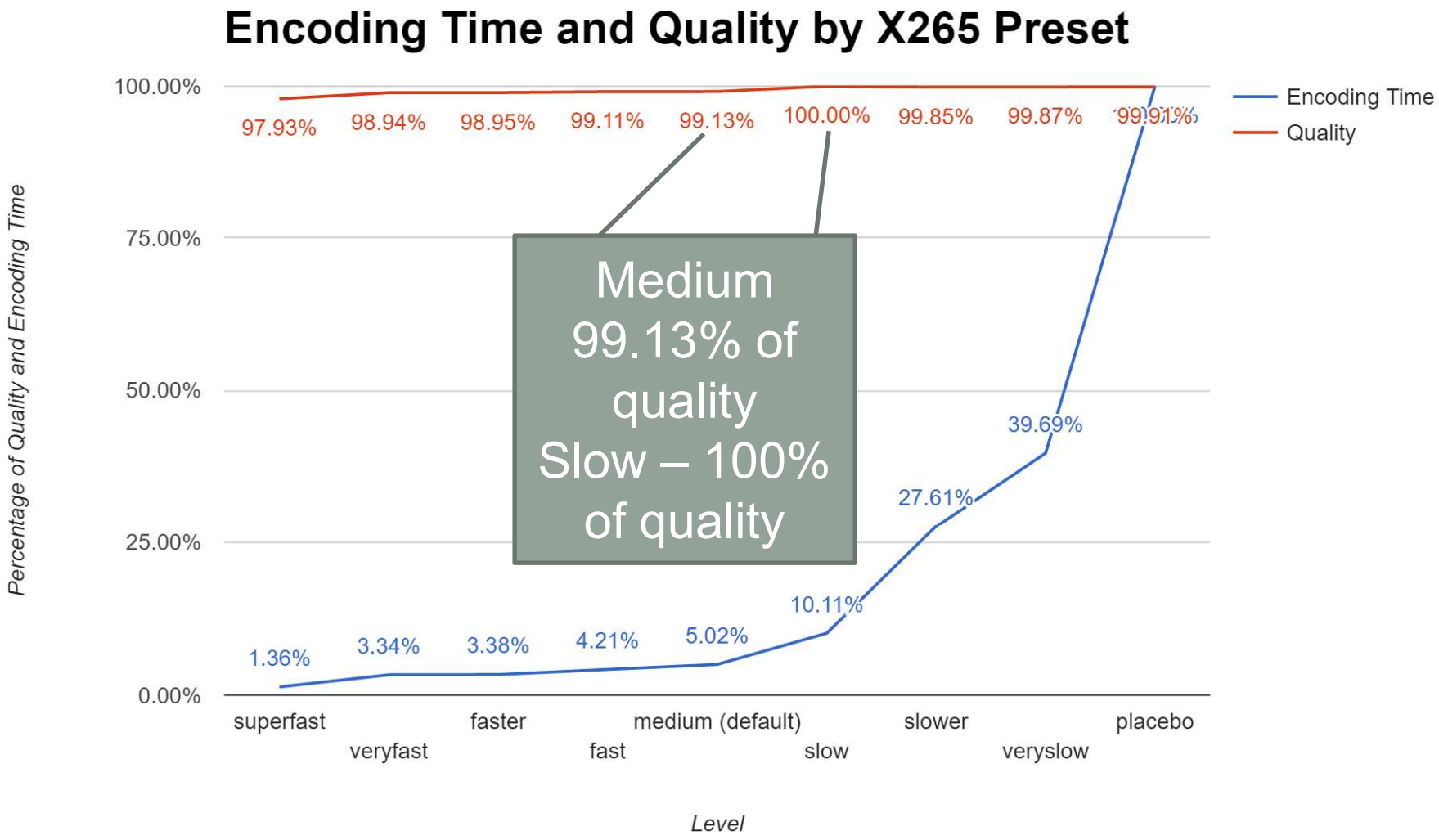
Choosing the Quality Setting

- Tried to find reasonable setting where MC/x265 performed similarly
- Performance tests
  - Encoded five files used SABET at all quality levels
  - x265 – encoded five files in parallel using Capella Systems Cambria Encoder at all quality levels
  - Found a duration that:
    - Produced reasonably quality with both codecs
    - Produced the output at about the same time
      - MainConcept – level 18
      - X265 – Medium preset
  - VP9 – Used most commercially reasonable setting (more later)
  - X264 – just used very slow preset which was much faster than all other tested technologies

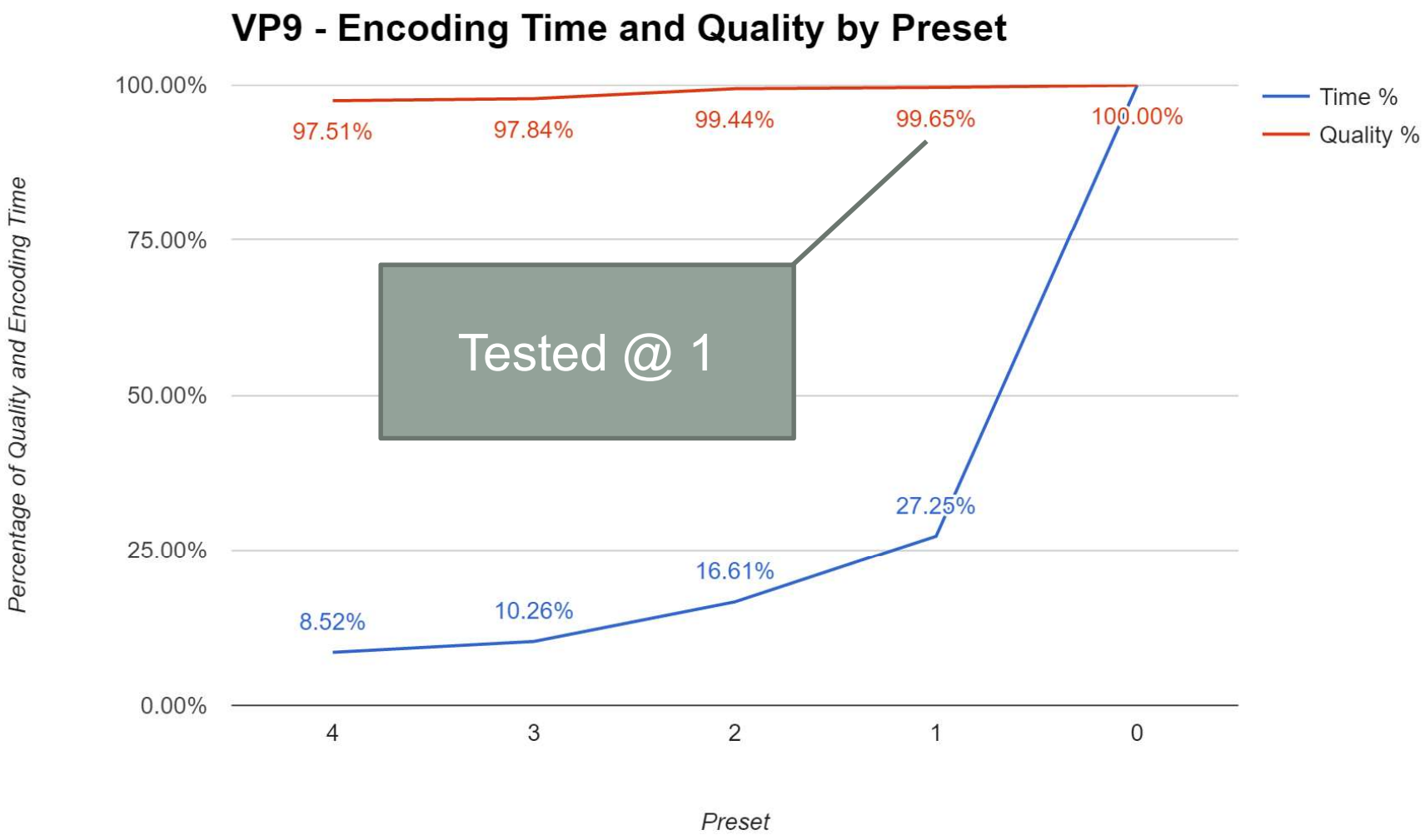
MainConcept – Times and Quality



x265 – Times and Quality



VP9 – Times and Quality



Measured Quality with VMAF

- Video Multimethod Assessment Fusion (VMAF)
  - Objective metric used by Netflix in their per-title optimization workflow
  - Replaced PSNR
  - Open source
  - Meld of multiple benchmarks

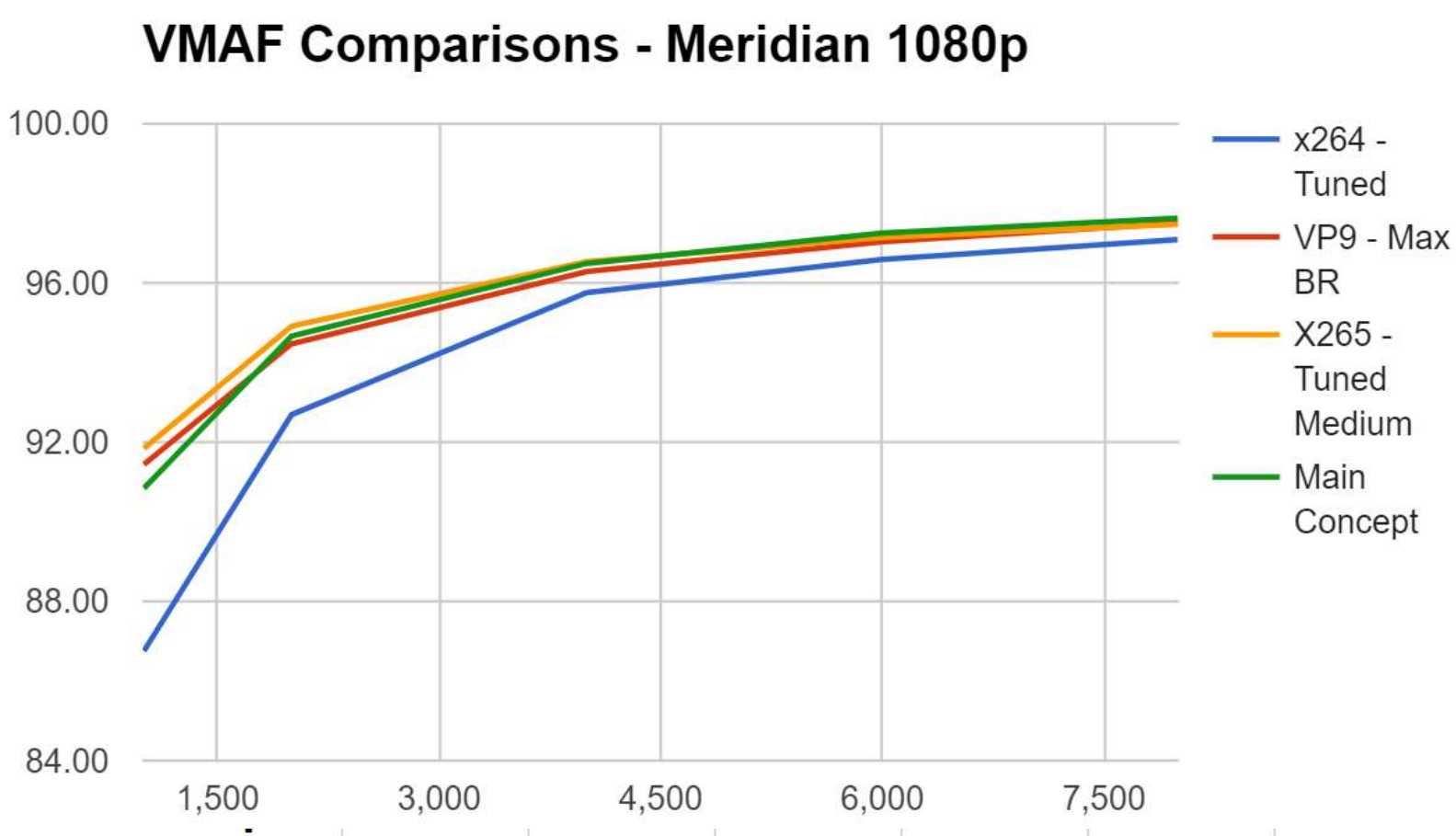
Measured Quality with VMAF

- Obtained scores from cloud encoding vendor www.hybrik.com
  - I performed all encodes
  - Uploaded source and encoded files to S3
  - Ran VMAF analysis only
  - Hybrik charges a monthly sum based on number of simultaneous machines (starting at $1,000 for 10 machines)
    - http://bit.ly/hybrik_cost
    - Recent winner of Streaming Media Best of NAB award

To Tune or Not to Tune (with VMAF)?

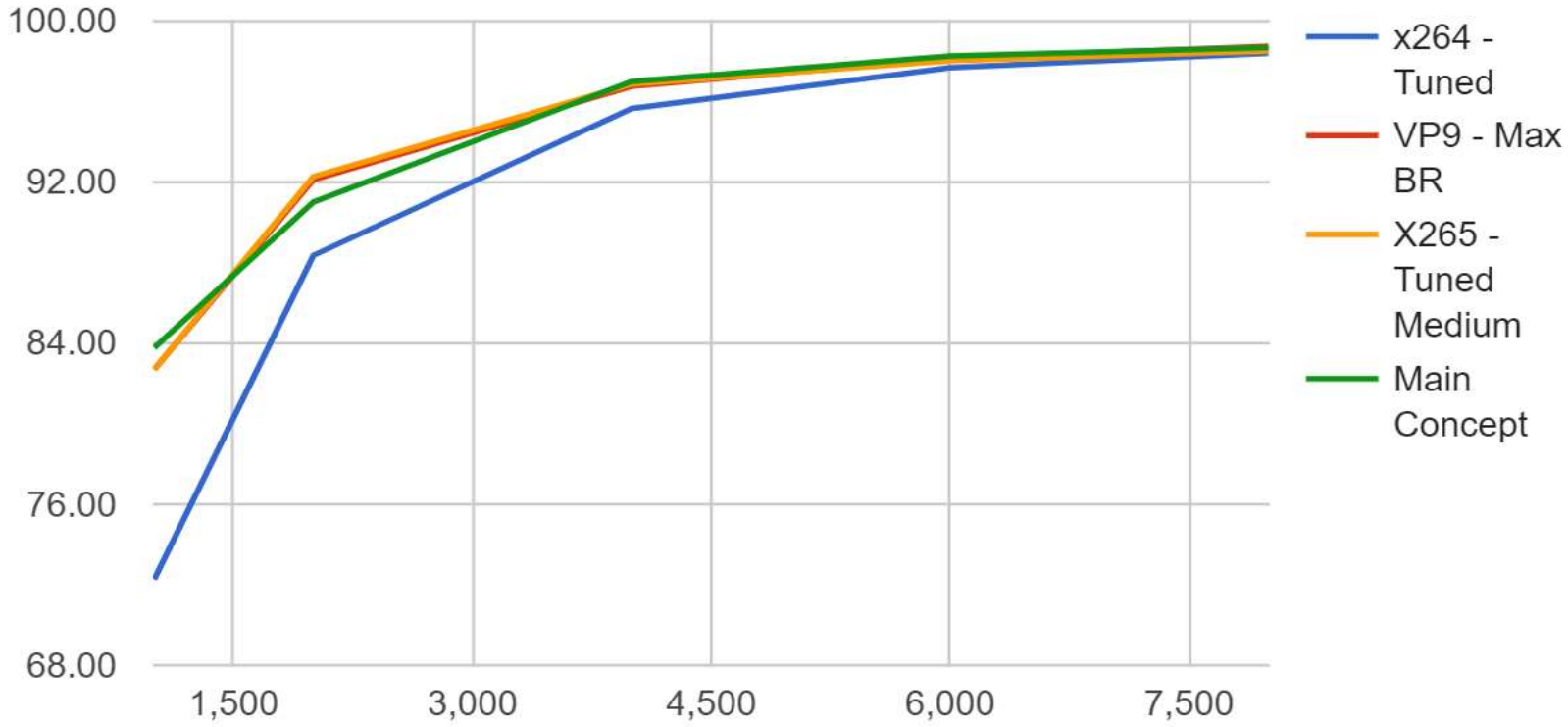
- Tom Vaughan - With VMAF I think you're safe to tune for visual quality (don't use --tune psnr or --tune ssim). VMAF isn't perfect, but it's the closest objective metric to subjective quality we've got at this point.
- Tuning produced better results in each comparison
- So, we tuned x265

	x265 Redo	x265 Tune	Delta
Meridian - 1080p			
4000	90.98	92.33	1.48%
8000	93.64	94.56	0.98%
12000	94.74	95.51	0.81%
16000	95.39	96.08	0.72%
20000	95.84	96.46	0.65%
Average	**94.12**	**94.99**	0.92%
Sintel - 1080p			
4000	87.79	89.21	1.61%
8000	94.32	94.90	0.62%
12000	96.39	96.81	0.43%
16000	97.35	97.60	0.25%
20000	97.98	98.24	0.27%
Average	**94.77**	**95.35**	**0.62%**
TOSN - 1080p			
4000	92.01	93.89	2.04%
8000	96.09	96.97	0.92%
12000	97.22	97.85	0.65%
16000	97.73	98.28	0.57%
20000	98.15	98.60	0.45%
Average	**96.24**	**97.12**	0.91%
Overall Average	**95.04**	**95.82**	



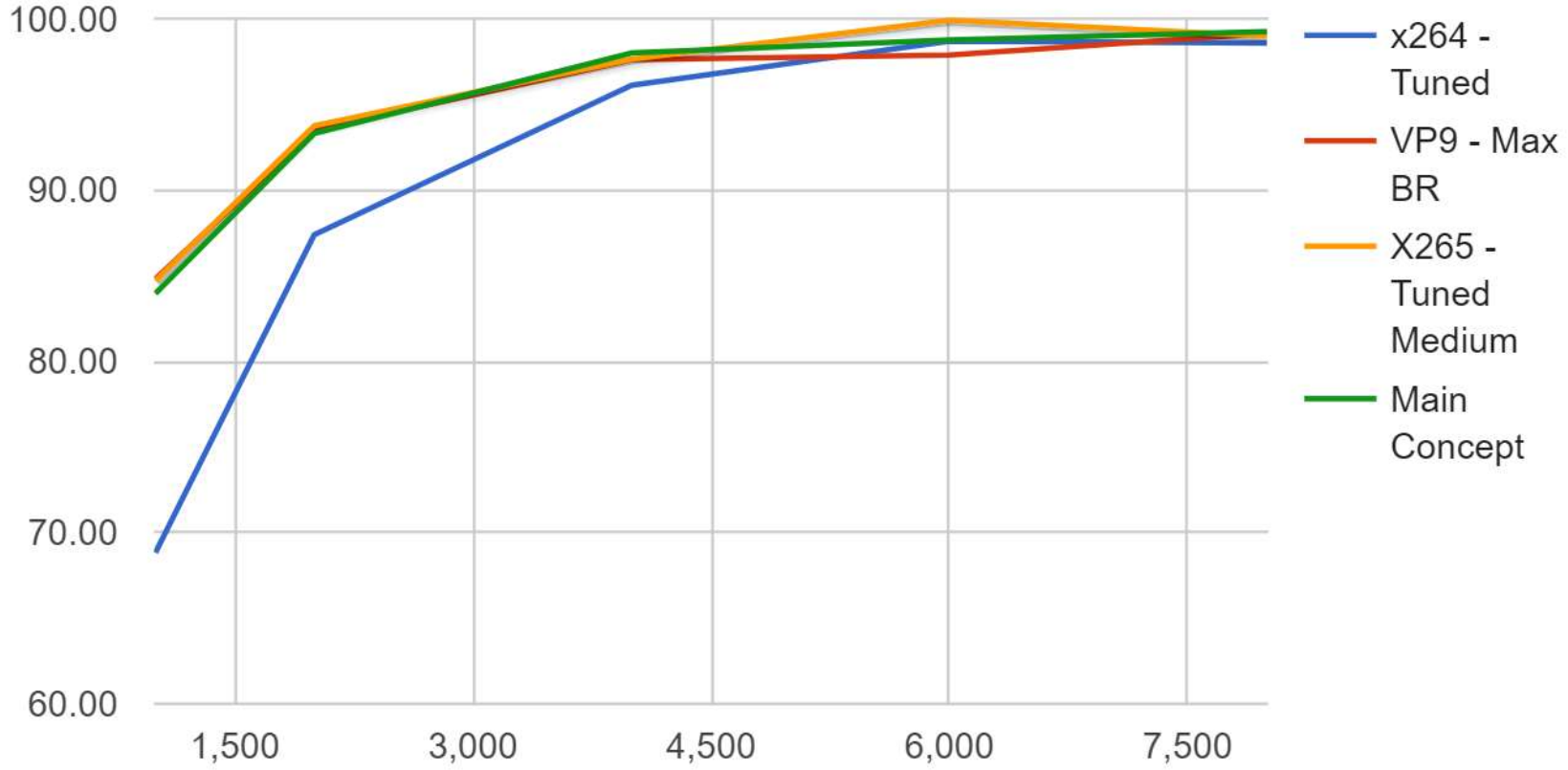
Meridian - 1080p	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Max Delta	MC/X265
1,000	86.75	91.44	91.84	90.84	5.87%	-1.10%
2,000	92.69	94.47	94.92	94.66	2.40%	-0.27%
4,000	95.76	96.29	96.54	96.49	0.81%	-0.04%
6,000	96.59	97.03	97.14	97.25	0.69%	0.11%
8,000	97.09	97.50	97.47	97.63	0.55%	0.15%
Average	**93.78**	**95.35**	**95.58**	**95.37**	**1.93%**	**-0.22%**

VMAF Comparisons - Sintel 1080p

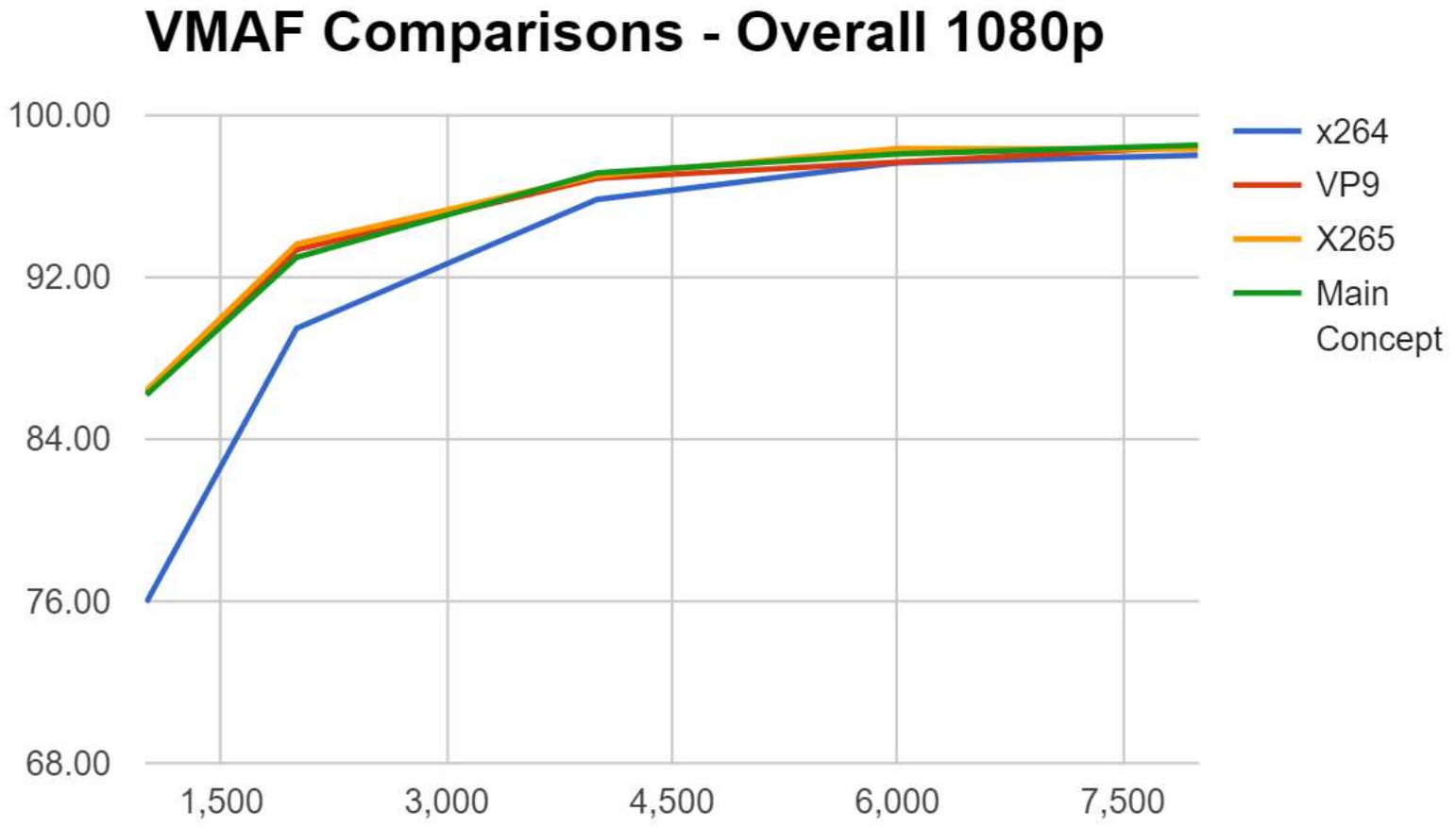


Sintel - 1080p	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Max Delta	MC/X265
1,000	72.27	82.70	82.71	83.78	15.92%	1.28%
2,000	88.37	92.13	92.28	91.01	4.43%	-1.39%
4,000	95.66	96.76	96.88	97.00	1.40%	0.13%
6,000	97.69	98.15	98.03	98.26	0.59%	0.24%
8,000	98.40	98.75	98.51	98.69	0.35%	0.19%
Average	**90.48**	**93.70**	**93.68**	**93.75**	**3.62%**	**0.07%**

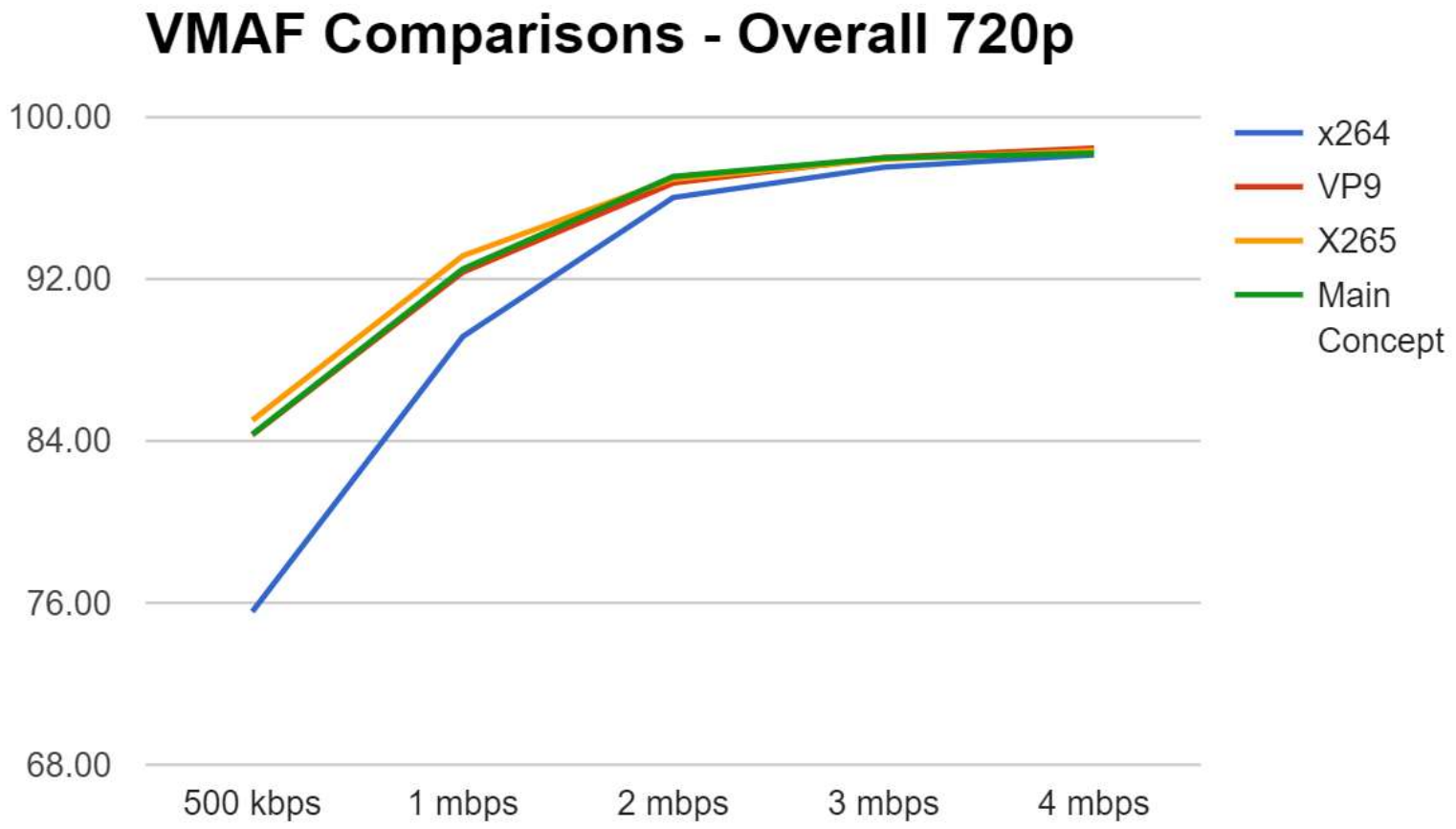
VMAF Comparisons - Tears of Steel 1080p



TOSN - 1080p	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Max Delta	MC/X265
1,000	68.83	84.80	84.70	83.95	23.21%	-0.89%
2,000	87.38	93.49	93.75	93.30	7.28%	-0.48%
4,000	96.12	97.61	97.64	98.00	1.96%	0.37%
6,000	98.68	97.88	99.91	98.77	2.08%	-1.16%
8,000	98.58	99.14	99.00	99.25	0.68%	0.25%
Average	**89.92**	**94.59**	**95.00**	**94.65**	**5.65%**	**-0.37%**

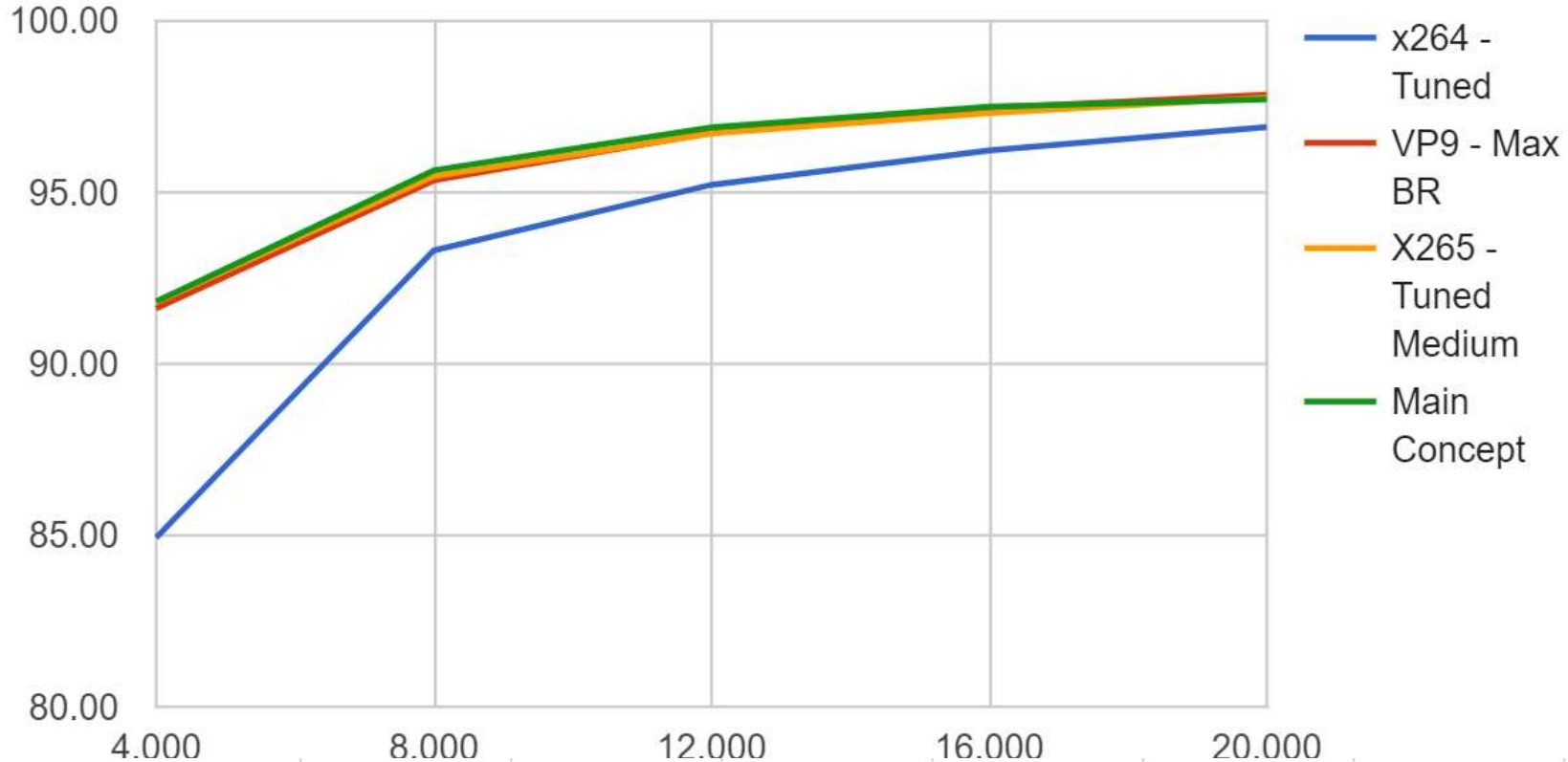


Overall Average - 1080p	x264	VP9	X265	Main Concept	Max Delta	MC/X265
1,000	75.95	86.31	86.42	86.19	13.78%	-0.26%
2,000	89.48	93.37	93.65	92.99	4.66%	-0.71%
4,000	95.85	96.89	97.02	97.16	1.37%	0.15%
6,000	97.65	97.69	98.36	98.09	0.73%	-0.27%
8,000	98.03	98.46	98.33	98.52	0.51%	0.20%
Average	**91.39**	**94.54**	**94.75**	**94.59**	**3.68%**	**-0.17%**



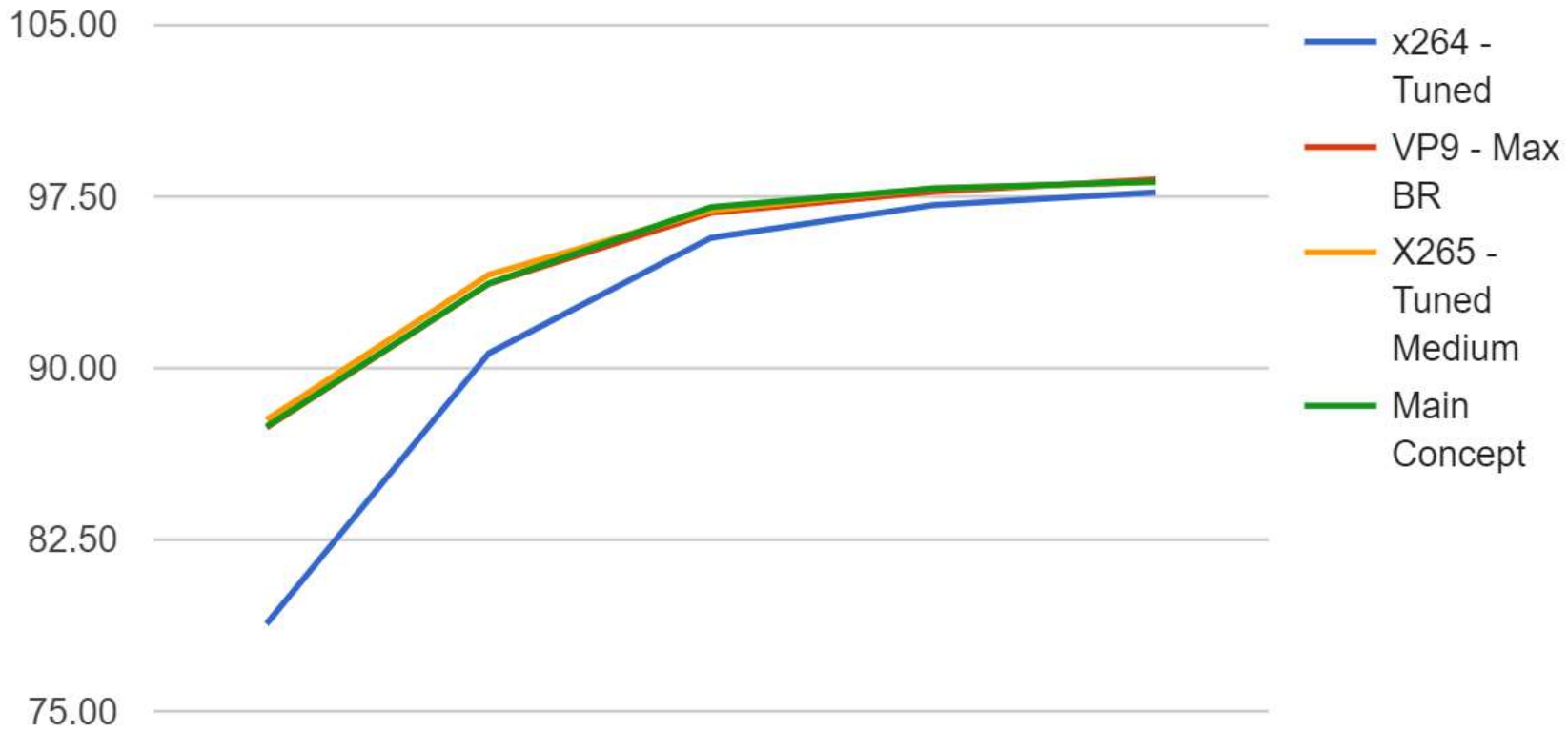
Overall - 720p	x264	VP9	X265	Main Concept	Max Delta	MC/X265
500 kbps	75.57	84.30	85.02	84.33	0.133	-0.82%
1 mbps	89.16	92.34	93.15	92.49	0.046	-0.72%
2 mbps	96.03	96.75	96.94	97.06	0.012	0.12%
3 mbps	97.53	98.00	97.93	97.99	0.005	0.06%
4 mbps	98.14	98.47	98.34	98.22	0.005	-0.13%
Average	**91.28**	**93.97**	**94.28**	**94.02**	**3.28%**	**-0.28%**

VMAF Comparisons - Overall 4K



Overall Average - 4K	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Max Delta	MC/X265
4,000	84.93	91.62	91.81	91.82	8.11%	0.02%
8,000	93.31	95.35	95.48	95.65	2.50%	0.18%
12,000	95.23	96.74	96.72	96.90	1.76%	0.19%
16,000	96.23	97.46	97.32	97.50	1.32%	0.18%
20,000	96.91	97.86	97.77	97.72	0.97%	-0.05%
Average	**93.32**	**95.81**	**95.82**	**95.92**	**2.78%**	**0.10%**

VMAF Comparisons - Overall Overall



Overall Overall	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Max Delta	MC/X265
	78.82	87.41	87.75	87.45	11.33%	-0.35%
	90.65	93.68	94.09	93.71	3.80%	-0.41%
	95.70	96.79	96.89	97.04	1.40%	0.15%
	97.13	97.72	97.87	97.86	0.76%	-0.01%
	97.69	98.26	98.15	98.15	0.58%	0.01%
Average	**92.00**	**94.77**	**94.95**	**94.84**	**3.21%**	**-0.11%**

Analysis

Overall Overall	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Max Delta	MC/X265
	78.82	87.41	87.75	87.45	11.33%	-0.35%
	90.65	93.68	94.09	93.71	3.80%	-0.41%
	95.70	96.79	96.89	97.04	1.40%	0.15%
	97.13	97.72	97.87	97.86	0.76%	-0.01%
	97.69	98.26	98.15	98.15	0.58%	0.01%
Average	**92.00**	**94.77**	**94.95**	**94.84**	**3.21%**	**-0.11%**

- Particularly at lower bitrates, both HEVC codecs and VP9 deliver substantially better performance than H.264
  - 11.33/3.80
- VP9 and both HEVC codecs produce very similar performance
- Choice between x265 and MC should be based on factors other than quality

Enter AV1

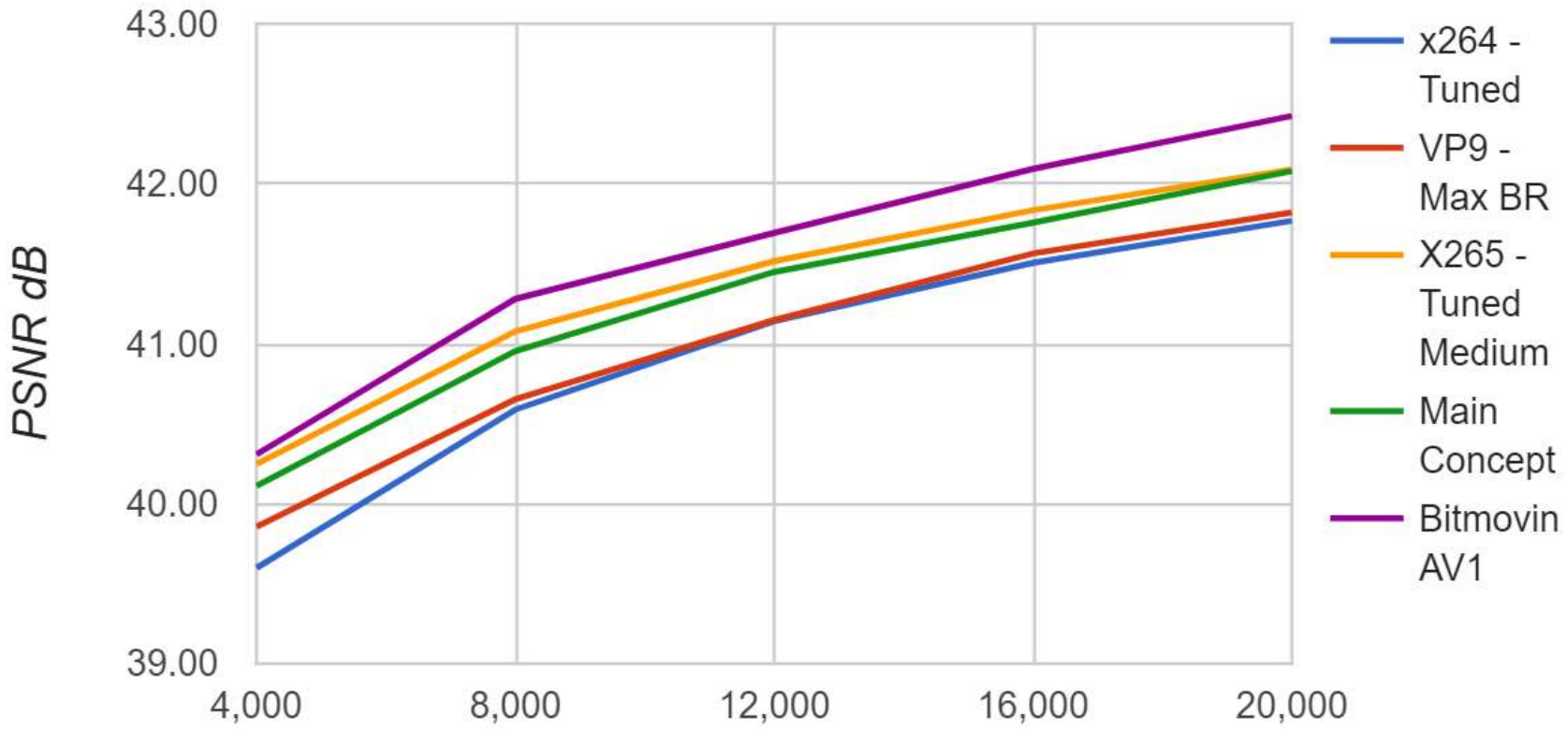
- Alliance for Open Media codec
  - Alliance formed by Google, Cisco, Mozilla, Amazon, Netflix, Microsoft and Intel
  - Goal was open source, royalty free codec
- First introduced in a live setting by cloud encoding/player/analytics vendor www.Bitmovin.com at NAB
  - Also won best of NAB award
- Very early days of codec
  - Unfair fight timewise – very long encode cycles
  - Did meet data rate requirements
  - Other basic parameters like GOP size
  - More technology demonstration than competition

Workflow

- Bitmovin encoded
- Sent WebM/DASH formatted files to me
- I decoded into Y4M
- Measured PSNR quality with Moscow University VQMT Tool
- Why no VMAF?
  - Hybrik didn’t have decoder
  - Y4M files too big to send

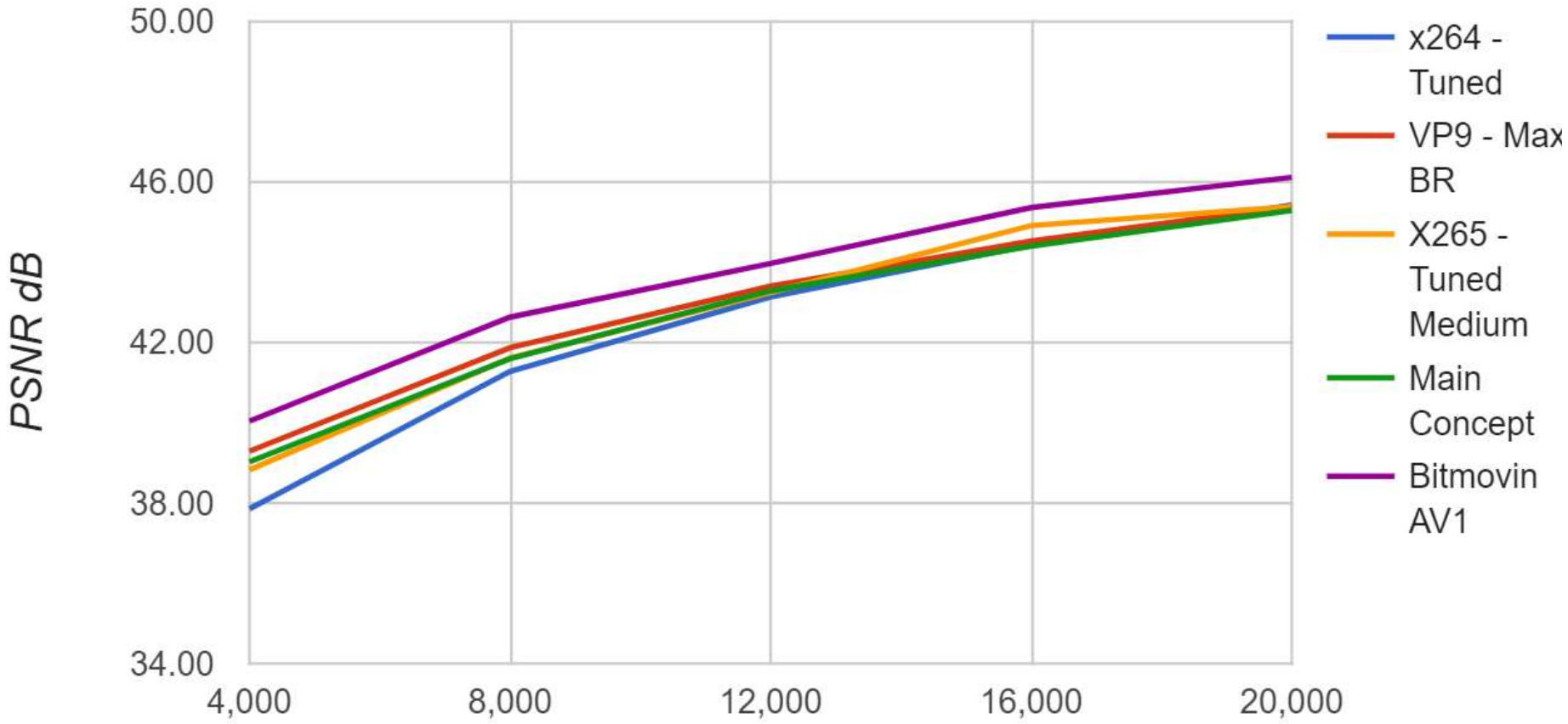
- Parameters used
  - quality = good
  - cpu-used = 0
  - threads = 0
  - lag-in-frames = 25
  - kf-max-dist = 60
  - kf-min-dist = 60
  - undershoot-pct = 5
  - overshoot-pct = 5
  - bitrate = 500

PSNR Comparisons - Meridian 4K



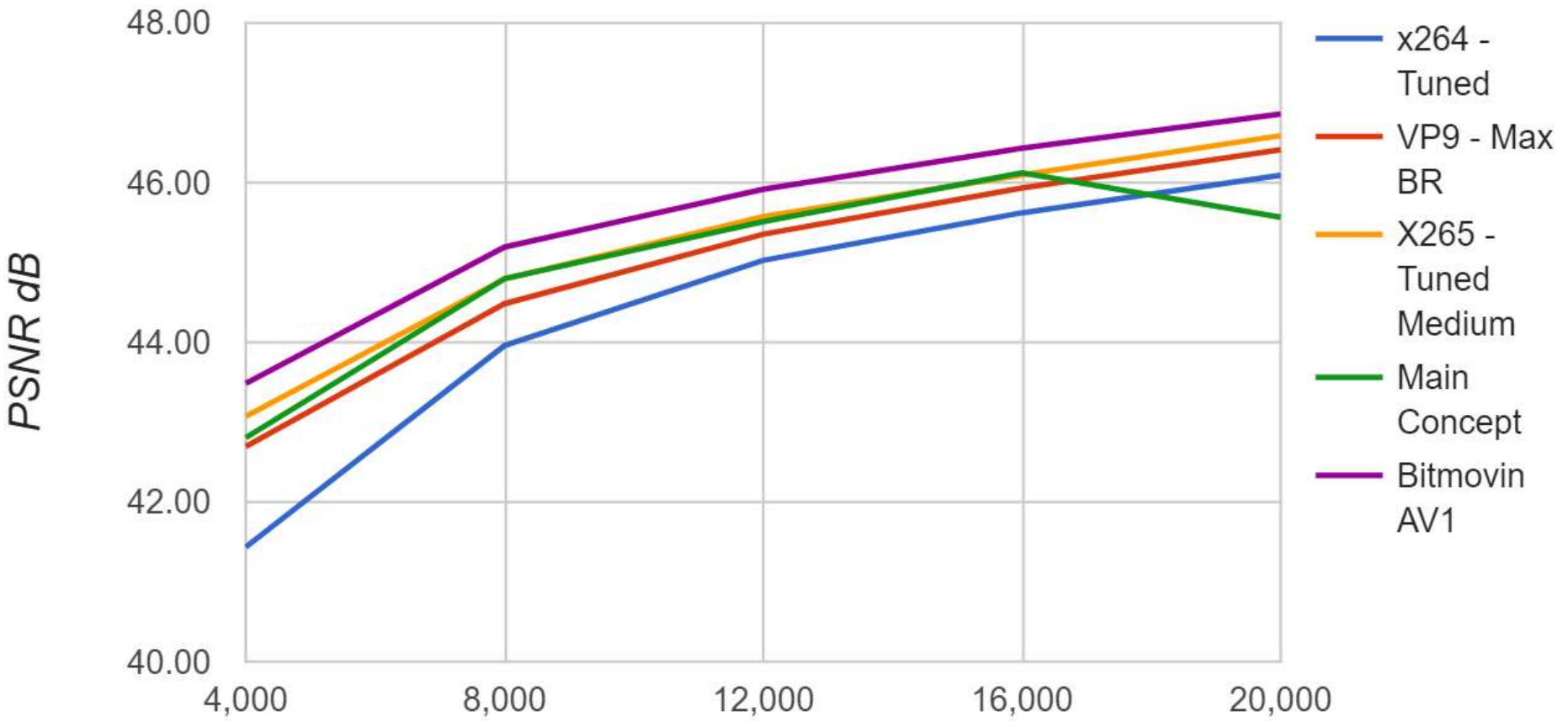
Meridian - 4K	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Bitmovin AV1	Max Delta	MC/X265
4,000	39.60	39.86	40.25	40.11	40.31	1.79%	-0.34%
8,000	40.59	40.65	41.08	40.95	41.28	1.70%	-0.30%
12,000	41.14	41.15	41.52	41.45	41.69	1.34%	-0.16%
16,000	41.50	41.56	41.83	41.76	42.09	1.42%	-0.19%
20,000	41.77	41.82	42.09	42.08	42.42	1.57%	-0.02%
Average	**40.92**	**41.01**	**41.35**	**41.27**	**41.56**	**1.56%**	**-0.20%**

PSNR Comparisons - Sintel 4K



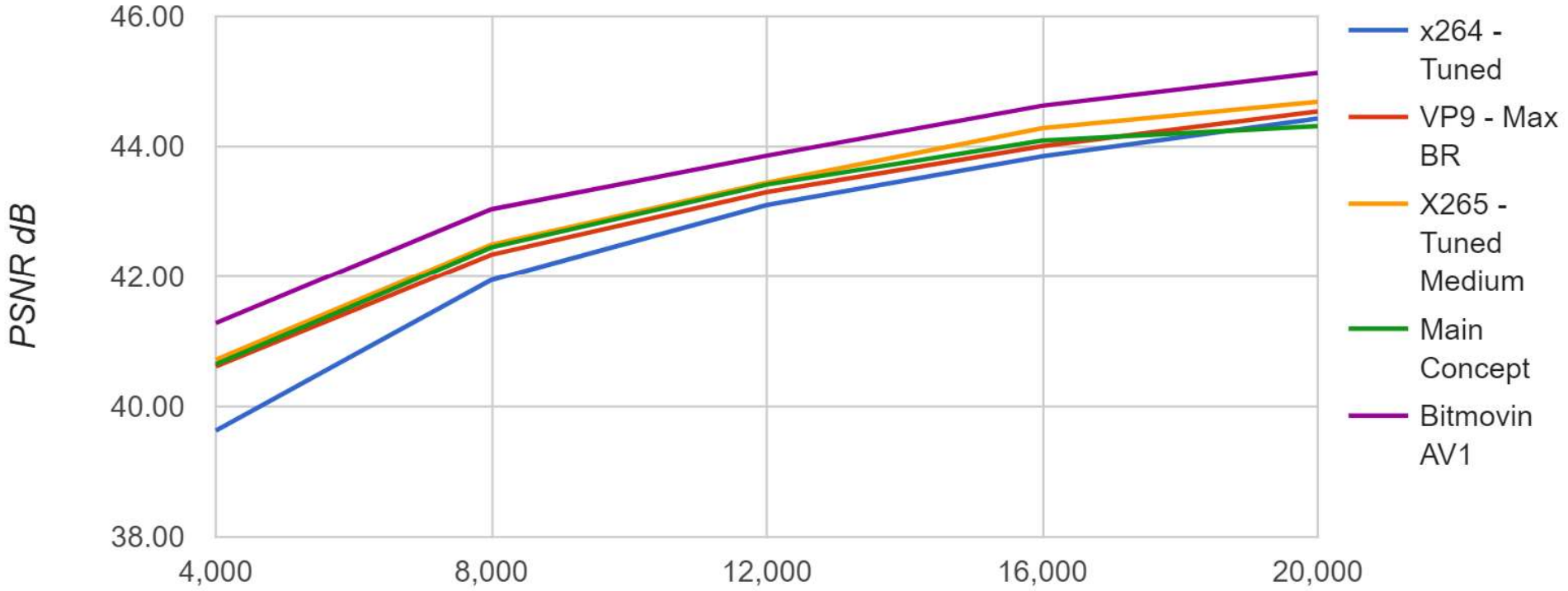
Sintel - 4K	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Bitmovin AV1	Max Delta	MC/X265
4,000	37.86	39.29	38.83	39.03	40.04	5.77%	0.51%
8,000	41.28	41.87	41.61	41.60	42.63	3.28%	-0.03%
12,000	43.14	43.40	43.25	43.29	43.97	1.93%	0.09%
16,000	44.43	44.53	44.92	44.40	45.36	2.16%	-1.16%
20,000	45.43	45.39	45.39	45.30	46.12	1.81%	-0.20%
Average	**42.43**	**42.90**	**42.80**	**42.72**	**43.63**	**2.83%**	**-0.18%**

PSNR Comparisons - TOSN 4K



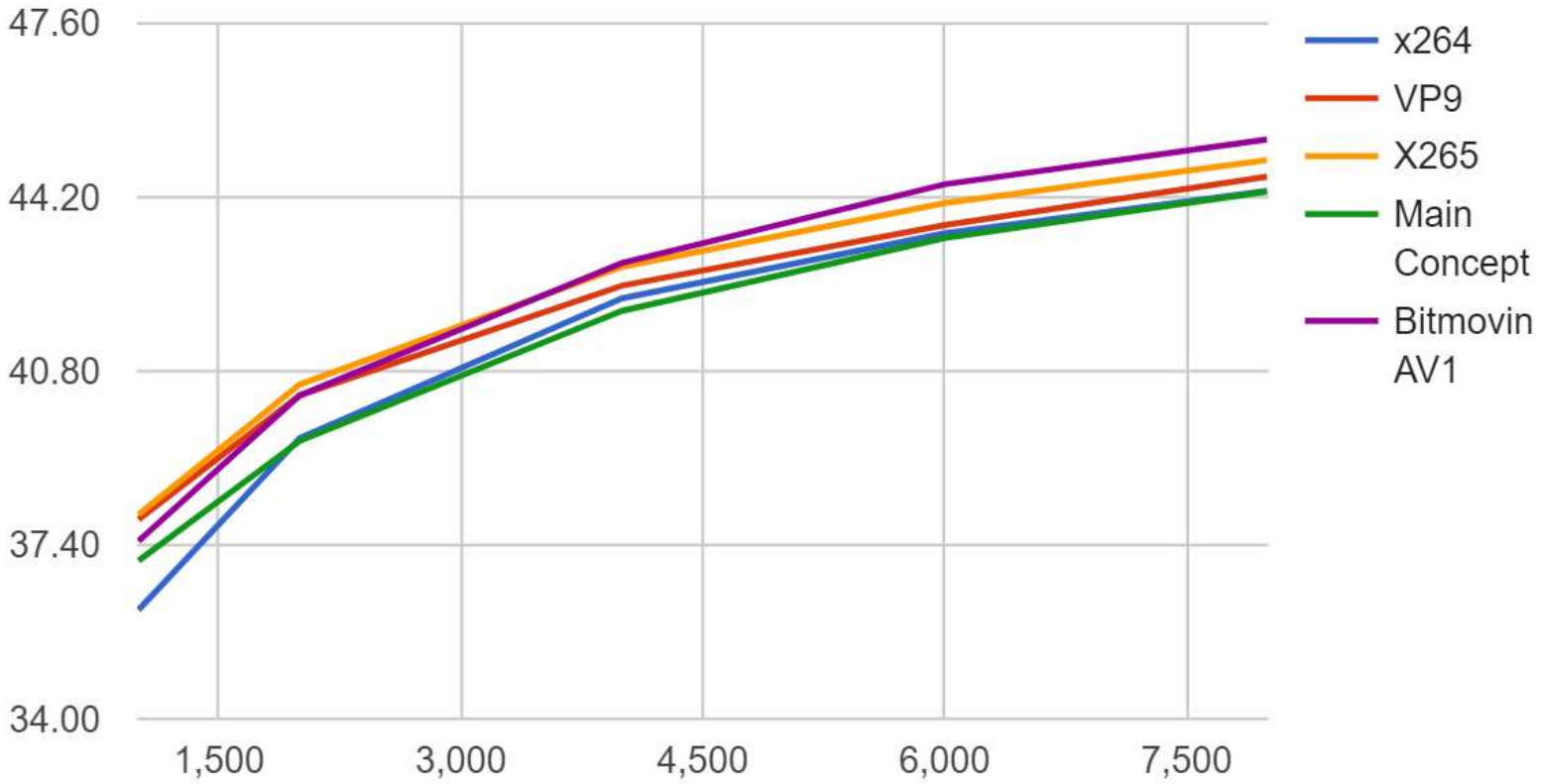
TOSN - 4K	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Bitmovin AV1	Max Delta	MC/X265
4,000	41.44	42.69	43.07	42.80	43.48	4.94%	-0.63%
8,000	43.96	44.48	44.79	44.80	45.19	2.81%	0.01%
12,000	45.03	45.35	45.57	45.51	45.92	1.98%	-0.14%
16,000	45.62	45.93	46.10	46.12	46.43	1.77%	0.06%
20,000	46.09	46.41	46.59	45.57	46.86	2.84%	-2.24%
Average	**44.43**	**44.98**	**45.23**	**44.96**	**45.58**	**2.59%**	**-0.59%**

PSNR Comparisons - Overall 4K



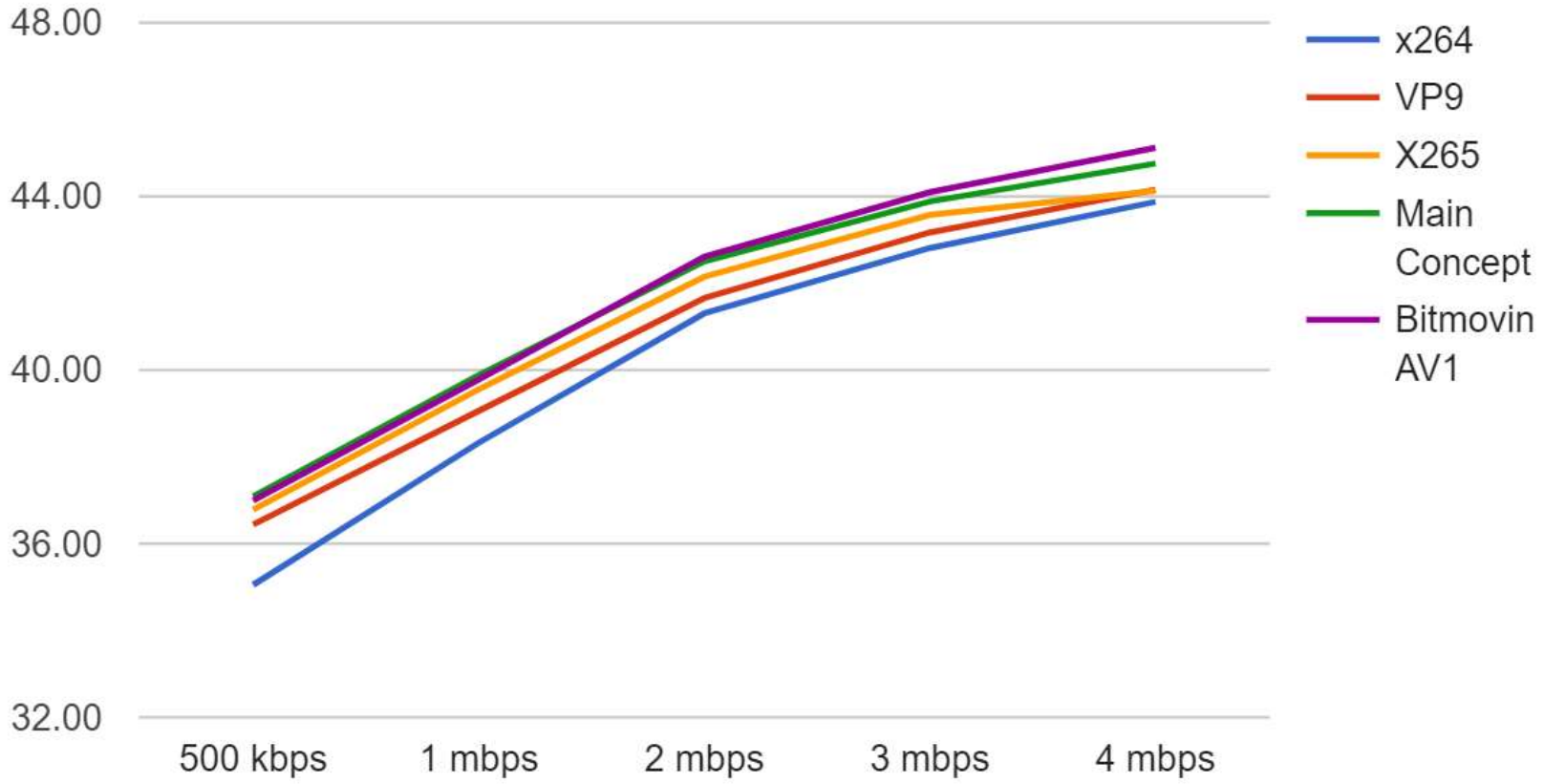
Overall Average - 4K	x264 -Tuned	VP9 - Max BR	X265 - Tuned Medium	Main Concept	Bitmovin AV1	Max Delta	MC/X265
4,000	39.63	40.61	40.72	40.65	41.28	4.16%	-0.17%
8,000	41.94	42.34	42.49	42.45	43.04	2.61%	-0.10%
12,000	43.10	43.30	43.45	43.42	43.86	1.76%	-0.07%
16,000	43.85	44.01	44.28	44.09	44.63	1.77%	-0.43%
20,000	44.43	44.54	44.69	44.31	45.13	1.85%	-0.84%
Average	**42.59**	**42.96**	**43.13**	**42.98**	**43.59**	**1.25%**	**-0.33%**

PSNR Comparisons - 1080p Overall



Overall Average - 1080p	x264	VP9	X265	Main Concept	Bitmovin AV1	Max Delta	MC/X265
1,000	36.13	37.90	38.00	37.09	37.48	5.16%	-2.44%
2,000	39.49	40.33	40.54	39.44	40.33	2.80%	-2.80%
4,000	42.22	42.47	42.84	41.98	42.92	2.23%	-2.05%
6,000	43.49	43.66	44.09	43.40	44.45	2.42%	-1.57%
8,000	44.33	44.61	44.93	44.31	45.33	2.30%	-1.39%
Average	**41.13**	**41.79**	**42.08**	**41.25**	**42.10**	**2.35%**	**-2.02%**

PSNR Comparisons - 720p Overall

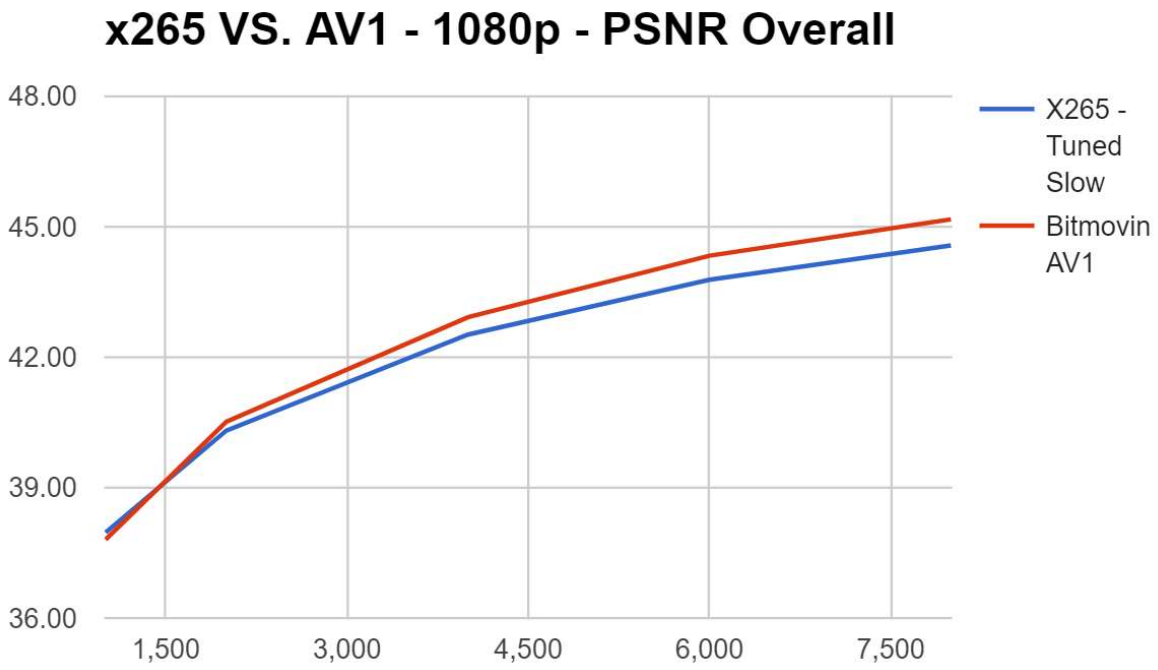


Overall - 720p	x264	VP9	X265	Main Concept	Bitmovin AV1	Max Delta	MC/X265
500 kbps	35.05	36.44	36.78	37.09	36.99	0.071	0.82%
1 mbps	38.33	39.06	39.56	39.88	39.78	0.054	0.82%
2 mbps	41.31	41.66	42.15	42.50	42.61	0.042	0.83%
3 mbps	42.81	43.17	43.57	43.88	44.10	0.038	0.71%
4 mbps	43.87	44.15	44.13	44.76	45.12	0.036	1.41%
Average	**40.27**	**40.90**	**41.24**	**41.62**	**41.72**	**3.59%**	**0.92%**

AV1 vs. x265 – Max Quality

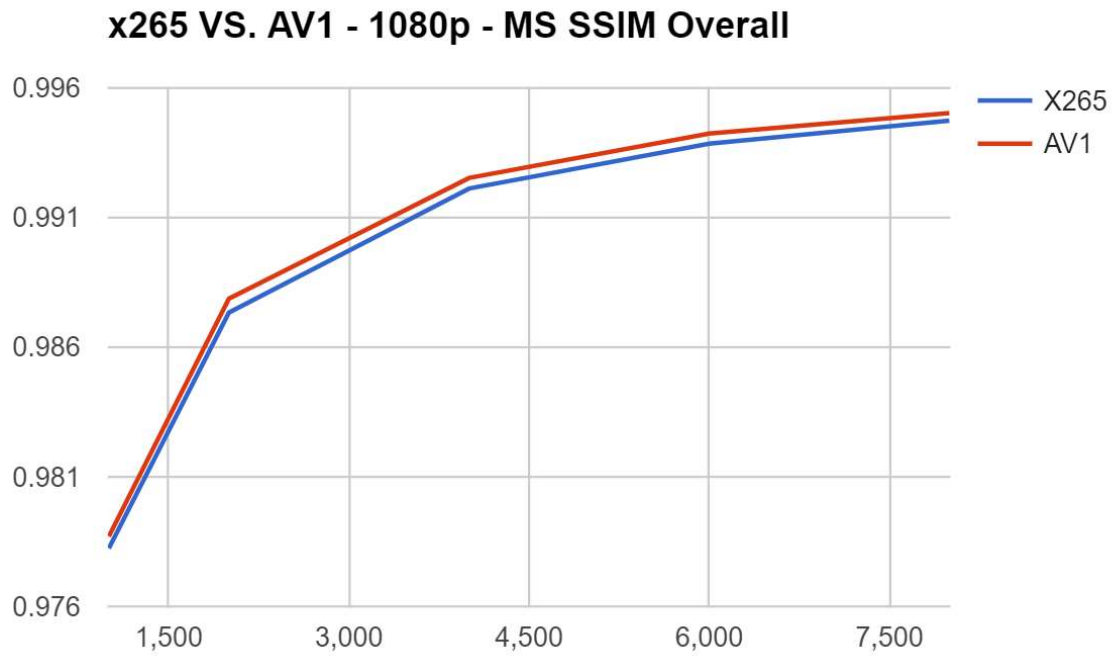
- Re-encode x265 using Slow preset
  - Saw before, this was max quality
- Substitute New test clip for TOSN
- All tests at 1080p
- Measure with PSNR and MS SSIM

PSNR Overall



Overall Average - 1080p	X265 - Tuned Slow	Bitmovin AV1	Max Delta
1,000	37.97	37.81	0.43%
2,000	40.31	40.52	0.50%
4,000	42.52	42.92	0.94%
6,000	43.78	44.33	1.26%
8,000	44.57	45.17	1.34%
Average	**41.83**	**42.15**	**0.76%**

MS SSIM Overall



Overall Average - 1080p	X265	AV1	Max Delta
1,000	0.978	0.979	0.05%
2,000	0.987	0.988	0.05%
4,000	0.992	0.993	0.04%
6,000	0.994	0.994	0.04%
8,000	0.995	0.995	0.03%
Average	**0.99**	**0.99**	**0.04%**

What's that Say About AV1?

- Bitstream scheduled to be frozen by end of 2017
- Other results
  - Google – AV1 is 35% more efficient than VP9
  - Netflix – AV1 is 20% more effective than VP9
- We're seeing less. Why?

Bitmovin Blog Post (bit.ly/AV1_BM)

- AV1 codebase is VP9/10
- There are 77 additional experimental coding tools that are under consideration.
- 8 are enabled by default (and by Bitmovin)
- Bitmovin is only AV1 system in production
  - More conservative than other experimental systems
  - Each user compiles their own version with selected experiments enabled
  - Don’t know how many coding tools included by Netflix and Google
  - Bottom line: results not necessarily inconsistent and will almost certainly improve at a rapid rate over time

Why So Different from Other Evals?

- Another report concluded “the H.265/MPEG-HEVC reference software implementation provides significant average bit-rate savings of 38.4% and 32.8% compared to AOM/AV1 and H.264/MPEG-AVC, respectively”
- Our results
  - Used commercially available HEVC encoder (not reference encoder)
  - Used a much more recent version of AV1 with more experiments compiled into the code
  - Used a version compiled by the company using it (Bitmovin), not the researcher
  - Used encoding parameters for VP9/AV1 (as well as x265/MC) that were approved by the vendors (not those compiled by the researcher)

Conclusions

- The qualitative difference between Main Concept and x265 was minimal, particularly with VMAF
- AV1 is at least as good as HEVC now, and will likely be quite a lot better when bitstream is frozen
  - Encoding times are still very inefficient
  - Encoding parameters still not explored
- VMAF seems to provide more accurate data than PSNR, particularly at aggressive data rates
  - More work needs to be done, particularly regarding the proper test settings for x264 and x265 (to tune, or not to tune).