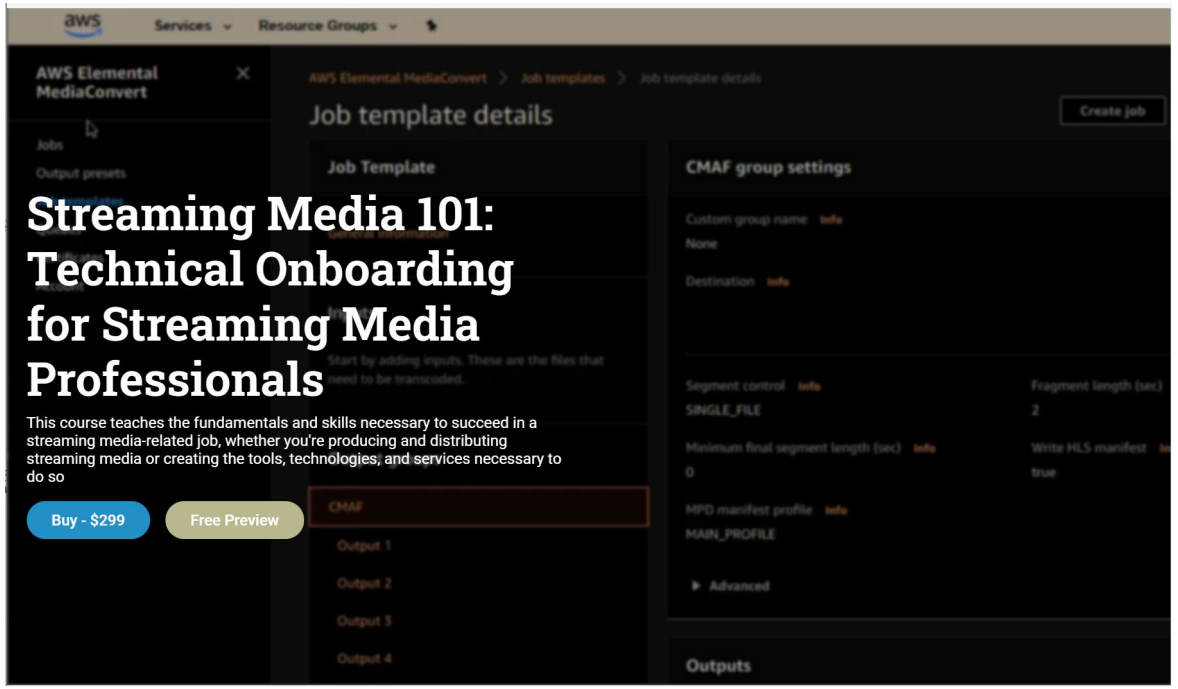
H.264 Encoding for Newbies and Optimizers

- Configuration basics
- Introduction to video quality metrics
- Encoding with H.264
  - Introduction
  - Compatibility related settings
  - Quality related settings
- Bitrate control techniques
- CRF and capped CRF
- Building your encoding ladder
- Per-title encoding

Promotional Note

- All lessons excerpted from my course Streaming Media 101: Technical Onboarding for Streaming Media Professionals
  - https://bit.ly/StreamingMedia101
- Adopted by Bitmovin and Dolby to onboard new staff in streaming-related job functions
  - https://bit.ly/dolby_sm101
  - http://bit.ly/bitmovin_sm101

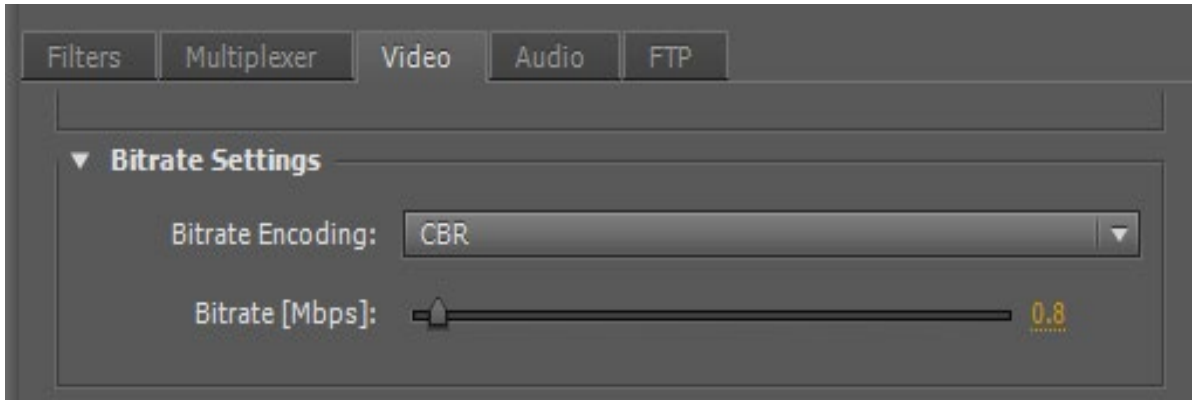


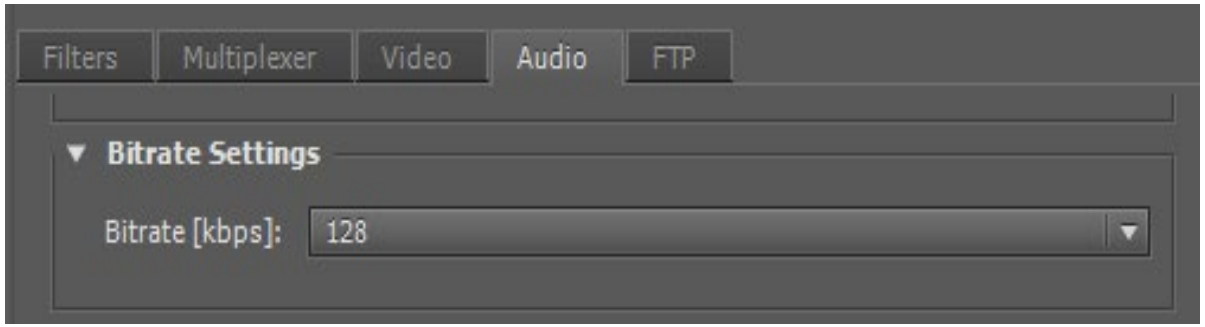
Lesson: Configuration Basics

- Encoding options you'll select for every file
  - Data rate
  - Resolution
  - Frame rate
- Encoding ladder

Configuration Basics – Data Rate

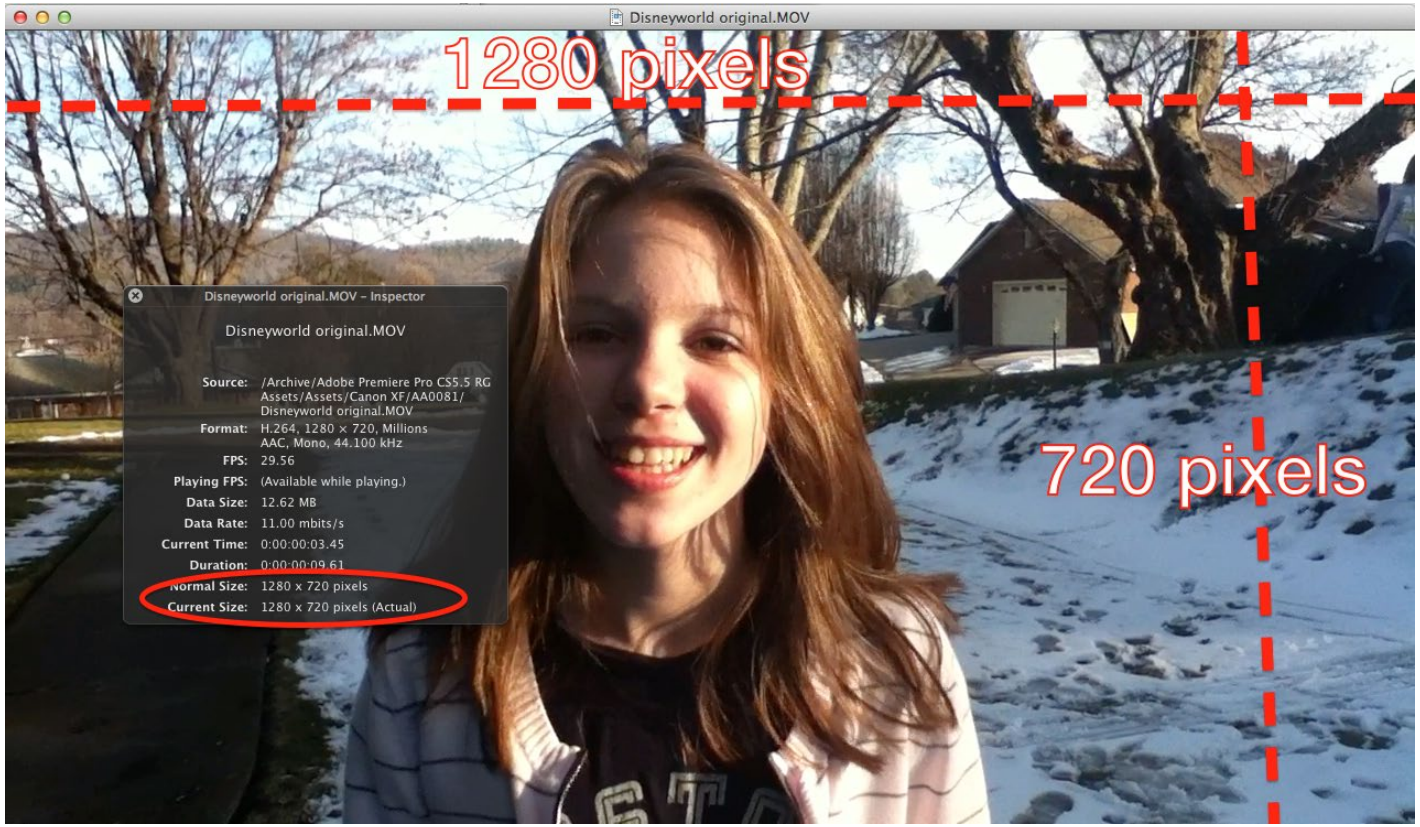
- You set data rate for video and audio for every file that you encode
- Video
  - Data rate is the most important factor in overall quality
  - The higher the data rate, the better the quality; but also harder to deliver
- Audio
  - For most audio, values beyond 128 kbps are a waste
  - Music videos and other high value productions may be the exception





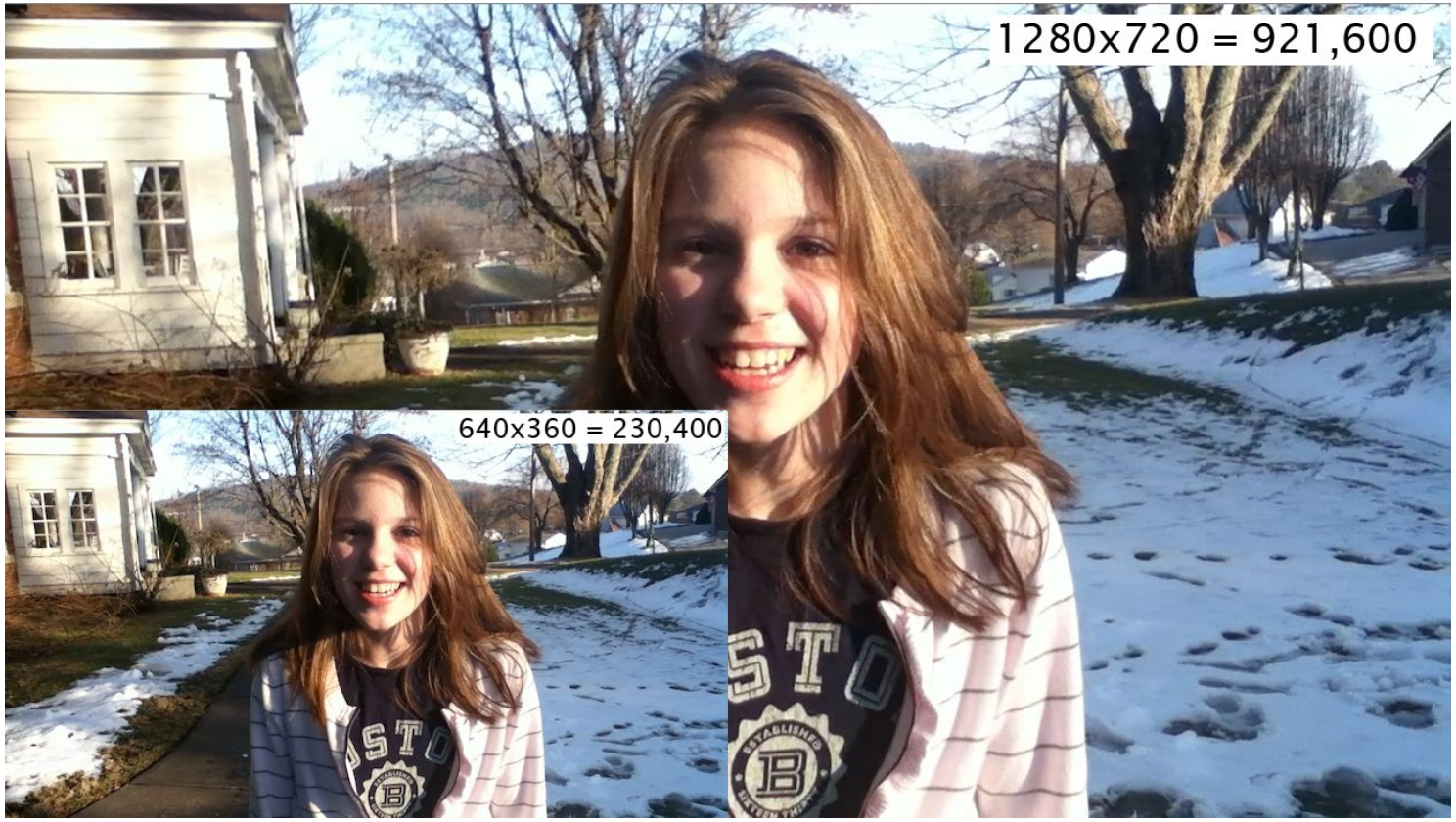
Configuration Basics – Video Resolution

- Width and height of video in a file
- Significant determinant of video quality
  - The more pixels, the harder a files is to compress
  - Fewer pixels, easier to compress



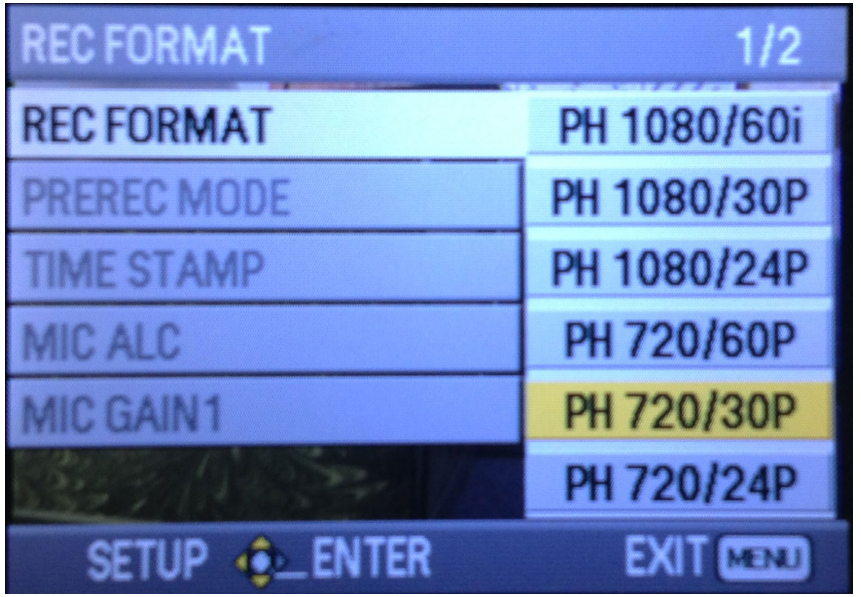
Configuration Basics – Video Resolution

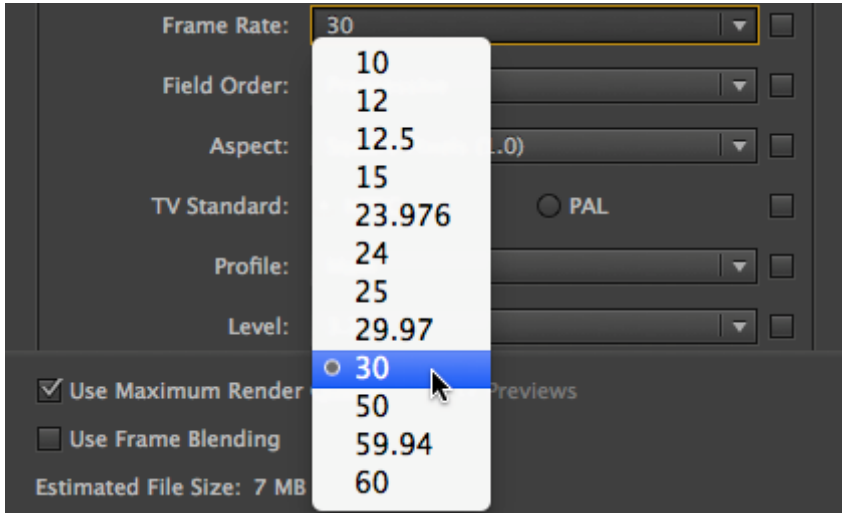
- That's why video files are often scaled down for streaming
  - 640x360 has ¼ the pixels of 1280x720



Configuration Basics – Frame Rate

- Frames per second in the file
- Set during recording (top)
- Usually maintained during streaming
  - Sometimes reduced for videos targeted for mobile devices





Encoding Ladder

Width	Height	Frame Rate	Video Bitrate	Audio Bitrate
416	234	12	200	64
480	270	15	400	64
640	360	29.97	600	64
640	360	29.97	1200	96
960	540	29.97	3500	96
1280	720	29.97	5000	128
1280	720	29.97	6500	128
1920	1080	29.97	8500	128

- Collection of files used for adaptive bitrate technologies like HLS/DASH
- Different resolutions, frame rates and A/V data rates to match devices and connection bandwidths

Lesson: Introduction to Objective Quality Metrics

- What they are
- Why we need them
- Meet VMAF
- Meet SSIMPLUS
- Meet PSNR

What Are Objective Quality Metrics

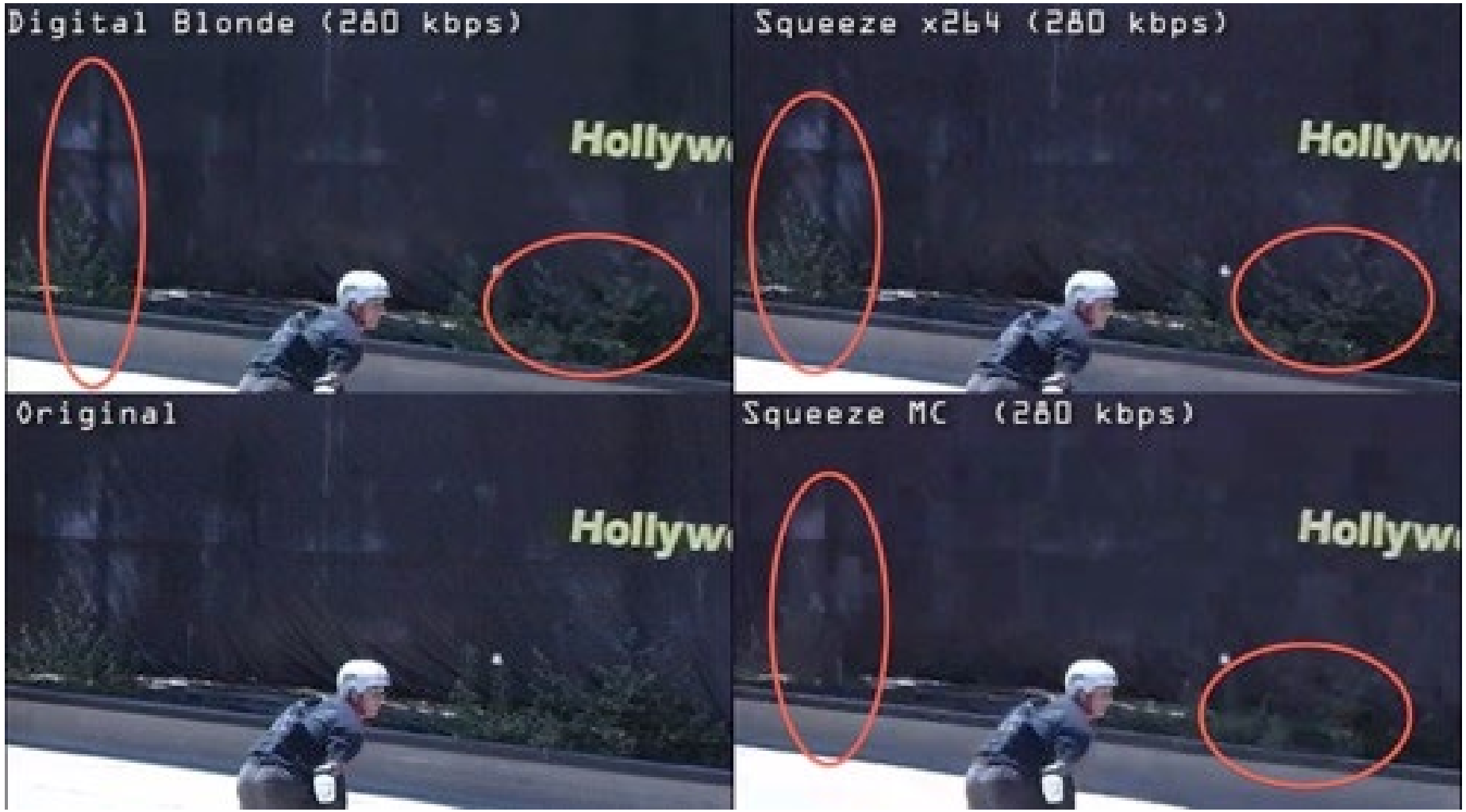
- Mathematical formulas that (attempt to) predict how human eyes would rate the videos
  - Faster and less expensive
  - Automatable
- Examples
  - Peak Signal to Noise Ratio (PSNR)
  - Structural Similarity Index (SSIM)
  - SSIMPLUS
  - VMAF (Video Multimethod Assessment Fusion)

Why Do We Need Them?

- So many encoding decisions
  - Data rate
  - Keyframe interval
  - B-frame interval
  - Bitrate control technique (VBR vs. CBR)
  - Choice of codec
  - Profile
  - Preset
- All have tradeoffs (quality vs. encoding time

- Objective quality metrics allow us to mathematically measure quality
- Uses
  - Drive many per-title encoding technologies (Netflix)
  - Useful to measure many critical encoding decisions
    - Many examples in this course

Took Me From Here



Time consuming and error prone
Subjective comparisons

To Here

VQM (lower is better)					
	Codec A	Codec B	Codec C	High > Low	Codec A > Codec B
Office 1	0.36	0.36	0.37	-3.54%	0.61%
Office 2	0.69	0.61	0.70	-13.51%	12.32%
Office 3	0.28	0.28	0.32	-14.74%	1.32%
Office 4	0.87	0.79	0.87	-9.63%	9.63%
Parking 1	0.68	0.61	0.74	-21.23%	10.90%
Parking 2	0.57	0.55	0.64	-15.47%	3.04%
Parking 3	1.86	1.58	1.76	-17.88%	17.88%
Parking 4	0.47	0.49	0.51	-8.86%	-3.81%
Retail 1	0.56	0.54	0.56	-4.27%	4.27%
Retail 2	0.68	0.66	0.69	-4.45%	3.39%
Retail 3	0.78	0.72	0.76	-8.64%	8.64%
Retail 4	0.73	0.67	0.88	-32.16%	8.52%
Traffic 1	0.55	0.50	0.58	-15.89%	9.14%
Traffic 2	0.34	0.32	0.38	-17.79%	6.39%
Traffic 3	0.52	0.49	0.55	-11.42%	5.29%
Traffic 4	0.68	0.61	0.66	-11.56%	11.56%
Total	**10.61**	**9.78**	**10.96**		
7.84%	Difference between Codec A and Codec B				
-3.34%	Difference between Codec A and Codec C				
-12.13%	Difference between Codec B and Codec C				
	0.61				
	Green equals best in category				
	Orange means worst in category				
	Difference greater than 7.5%				

Statistically meaningful comparisons

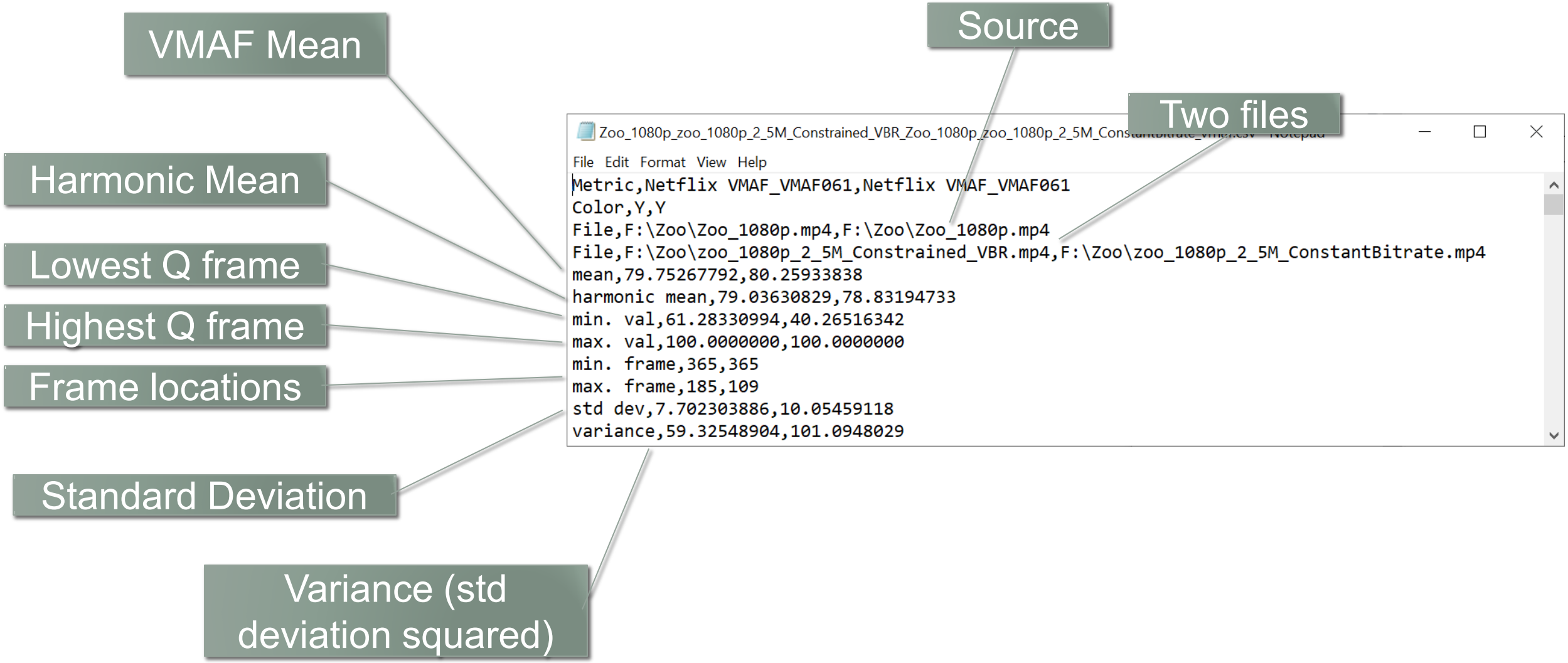
With Objective Quality Metrics You Get

- More data
  - Can run many more tests in much less time
- Better data
  - Mathematical models can detect smaller changes than your eye can easily discern

What is VMAF?

- Metric invented by Netflix
- Combines four existing metrics
- Fused using a Support Vector Machine (SVM)-based regression to a single output score ranging from 0–100 per video frame
  - 100 being identical to the reference video
  - Frame values are averaged to compute a single score

Numerical Results from CSV



VMAF Models

- Original (Default) model
  - Assumed that viewers watch a 1080p display with the viewing distance of 3x the screen height (3H).
- Phone model
  - Assume viewers watch on a mobile phone
- 4K Model
  - Video displayed on a 4K TV and viewed from a distance of 1.5H

1080p display

Mobile Phone

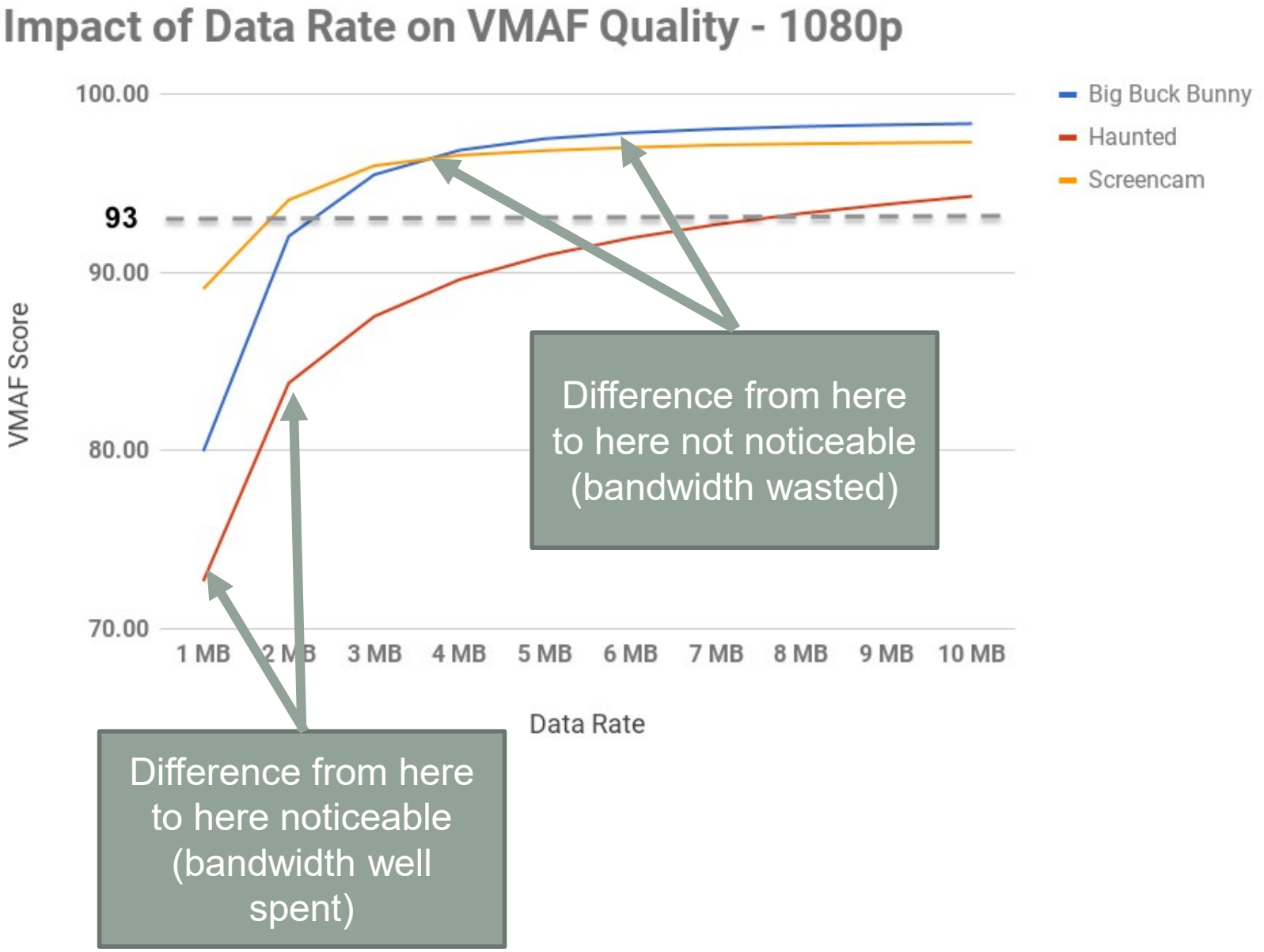
4K display

Working with VMAF – 93 is the Top Level Target

- Real Networks White Paper - VMAF Reproducibility: Validating a Perceptual Practical Video Quality Metric
  - 4K 2D videos
- ***VMAF score of about 93 … is either indistinguishable from original or with noticeable but not annoying distortion.***
  - http://bit.ly/vrqm_5

Working With VMAF

- Range – 0 – 100
- Top rung target – typically 93 – 95
  - Higher is a waste
- Scores map to subjective
  - 0-20 bad - 20 – 40 poor
  - 40 – 60 fair - 60 – 80 good
  - 80 – 100 excellent
- 6 VMAF points = Just noticeable difference

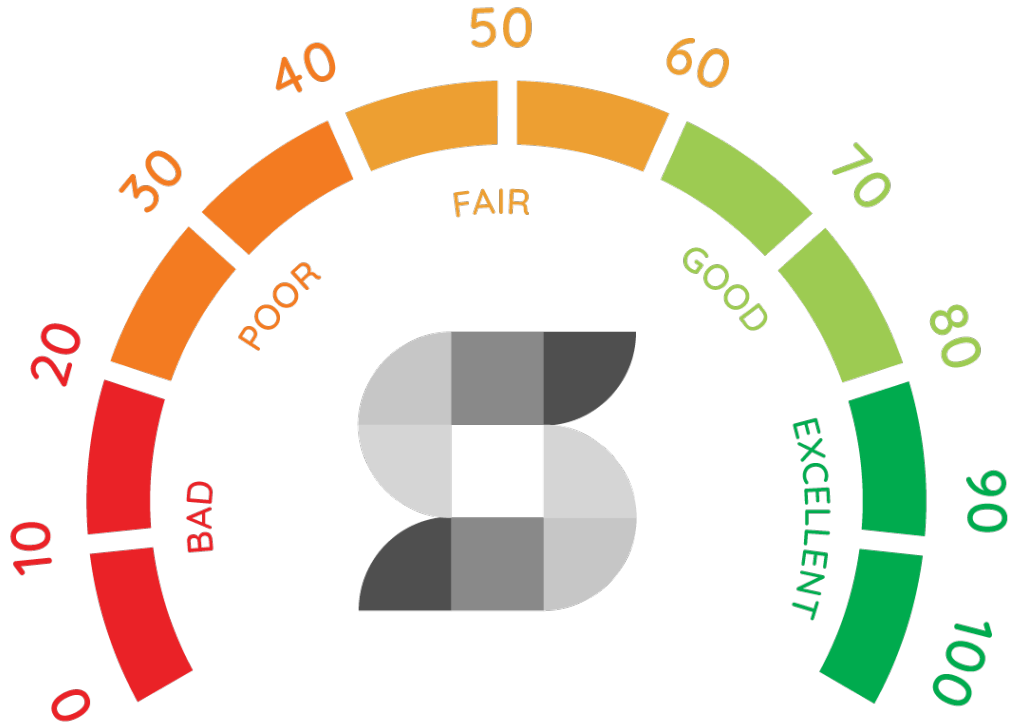


What is SSIMPLUS?

- Invented by inventor of SSIM
- Advancement of SSIM, extended to target video applications
- Strong correlation with subjective evaluations
- Scores map to easily understandable subjective ratings
- Supports multiple resolutions
- Supports multiple frame rates
- Supports some HDR formats
- Includes multiple device profiles
- Very fast

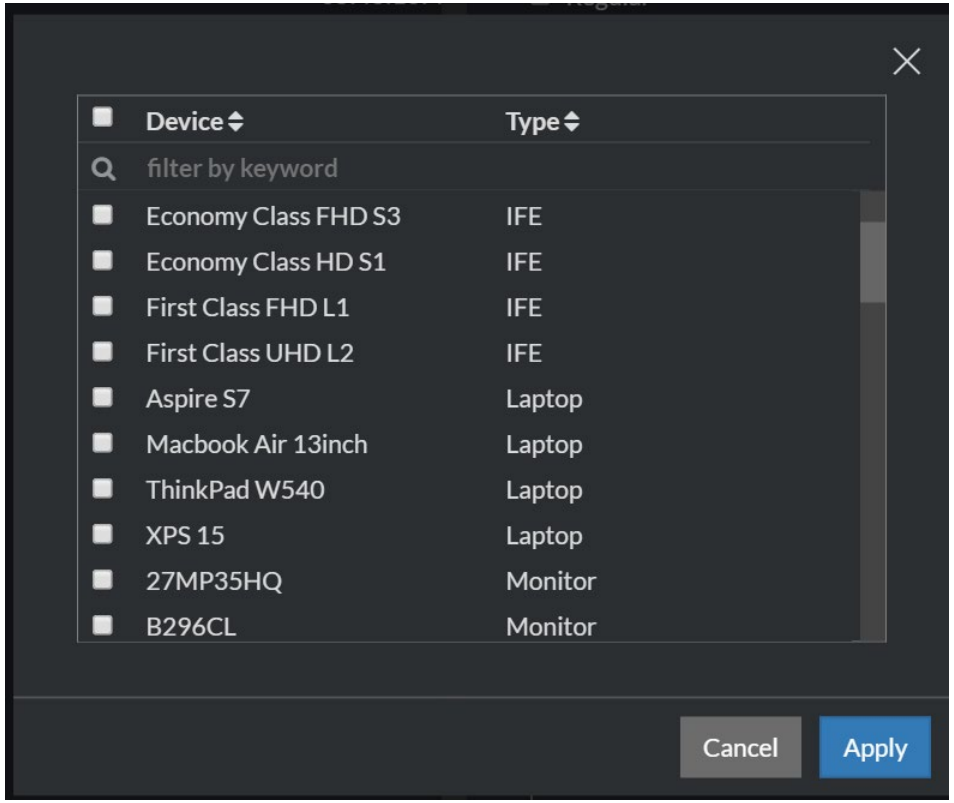
Working With SSIMPLUS

- SSIMPLUS scores easily map to subjective ratings
  - 0-20 bad
  - 20 – 40 poor
  - 40 – 60 fair
  - 60 – 80 good
  - 80 – 100 excellent



SSIMPLUS Device Models

- All scores reported for generic device plus many specific devices
  - Airline LCD panels
  - Smartphones
  - Tablets
  - Computer monitors
  - 1080p and 4K television sets
- Can assess quality on any and all devices relevant to your business
- Can customize encoding ladders by device



What About PSNR?

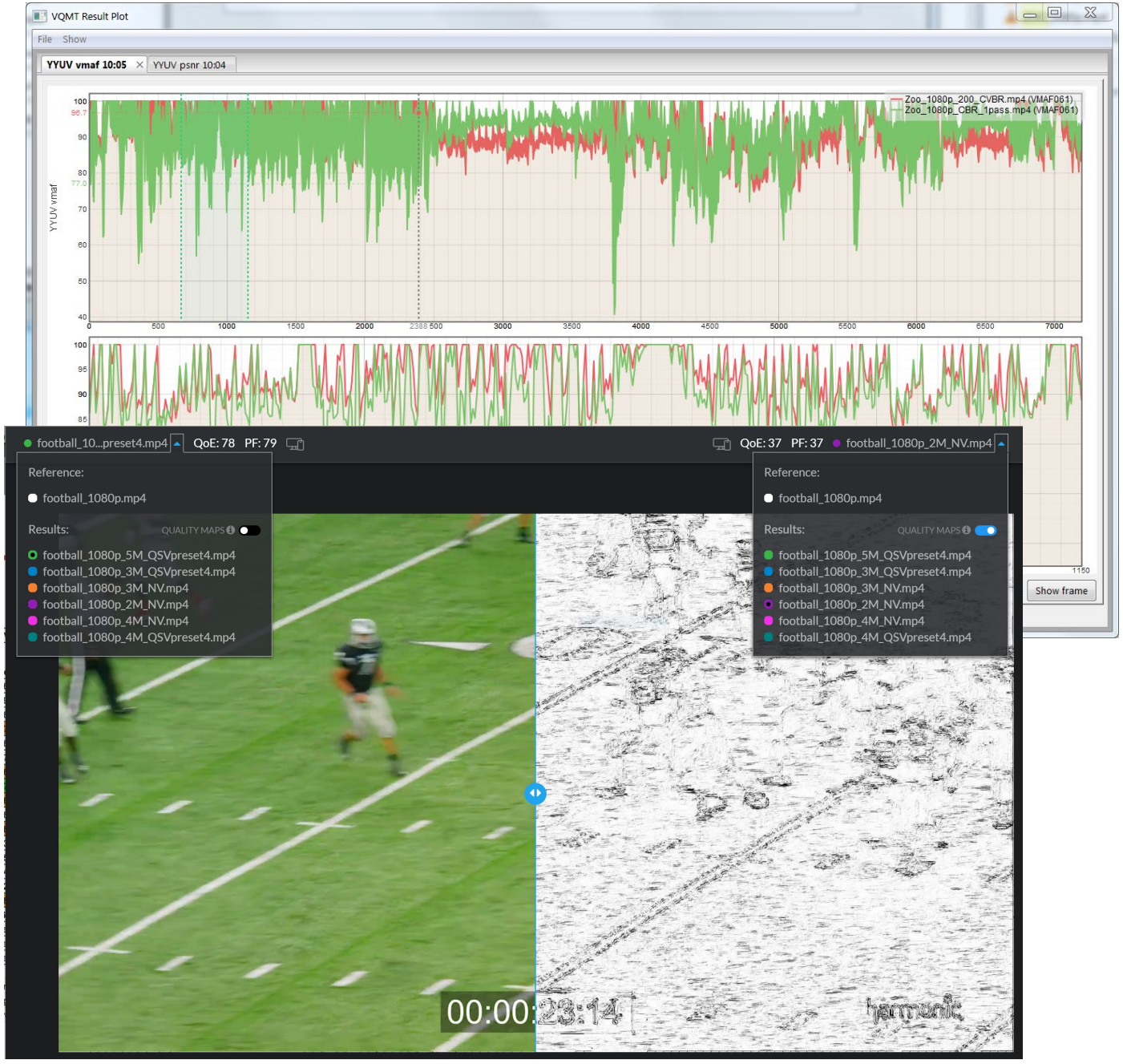
- Older, better known metric
- Still cited by Facebook and Netflix (and many researchers) in their codec-related conclusions
- Not as useful in the context of a complete encoding ladder
  - 1080p scores are reasonably accurate
  - Lower resolutions not so much

Computing Metrics

- VMAF
  - Moscow State University VQMT (~$1,000) - http://bit.ly/msu_vqmt
  - FFmpeg - Lesson of the Week: Computing VMAF with FFmpeg on Windows - http://bit.ly/vmaf_ffmpeg
  - Installing and Using Netflix VMAF-Master - http://bit.ly/VMaster
- SSIMPLUS
  - SSIMWAVE VOD Monitor (and other products) www.ssimwave.com
- PSNR
  - FFmpeg

Need a Visual Tool

- The best tools let you:
  - Identify potential problem frames
  - Visualize them
- FFmpeg great for summary numbers, but not great for true analysis



In this Session

- Mostly VMAF
  - Open-source, accessible to all
  - Consulting practice includes SSIMPLUS as well

Lessons: Encoding with H.264

- About H.264
- Compatibility-related settings
  - Profiles
  - Levels
  - Entropy encoding
- Performance-related settings
  - Choosing an x264 preset

What is H.264?

- Adopted by ISO and ITU
  - Telephony/cellular
  - TV - consumer electronics
  - Computer electronics
- Two names
  - MPEG/ISO – AVC (MPEG-4 part 10)
  - ITU – H.264

	ITU – International Telecommunications Union Telephone, Radio, TV	ISO – International Standardization Organization Photography, Computer, Consumer Electronics
1984	H.120	
1990	H.261 – Video Conferencing	
1993		MPEG-1 – Video CD
1994	(H.262)	MPEG-2 – Digital Cable and Satellite TV
1995	H.263 – Improved Video Conferencing	
1997		ATSC – U.S. HDTV
1999		MPEG-4
2002	AVC (H.264)	AVC (MPEG-4 Part 10)

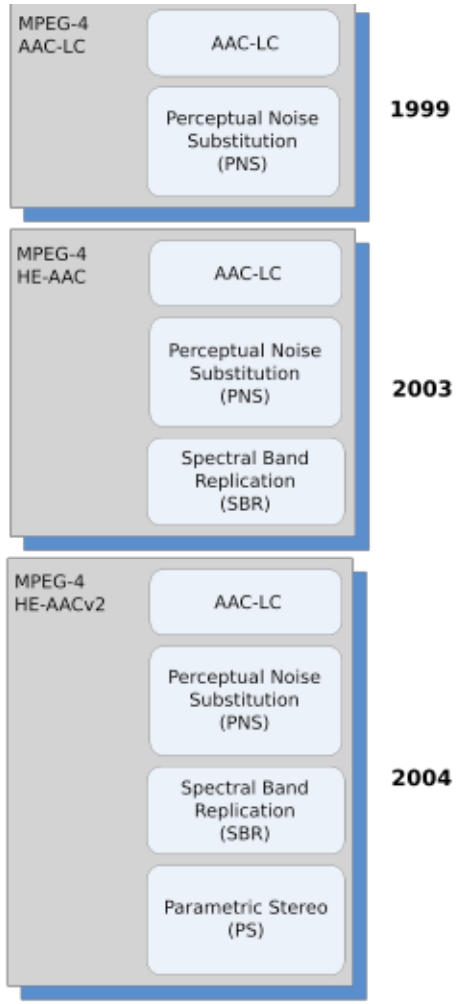
There are Multiple H.264 Codecs

- Because H.264 is a standard, there are many compliant codecs
  - FFmpeg uses the x264 H.264 codec (which is the best known and most widely used)
  - Elemental has its own
  - Adobe uses MainConcept
  - Beamr has their own (for licensing)

Critical H.264 Encoding Parameters

- Some parameters apply to all H.264 codecs
  - Profiles, levels, entropy coding
  - No matter which H.264 codec you work with, you must set these
- Some are codec specific
  - Presets (x264's way of balancing encoding time and quality)
  - Other codecs may have a similar control, but only x264 (and x265) has presets called slow, very slow, placebo etc.

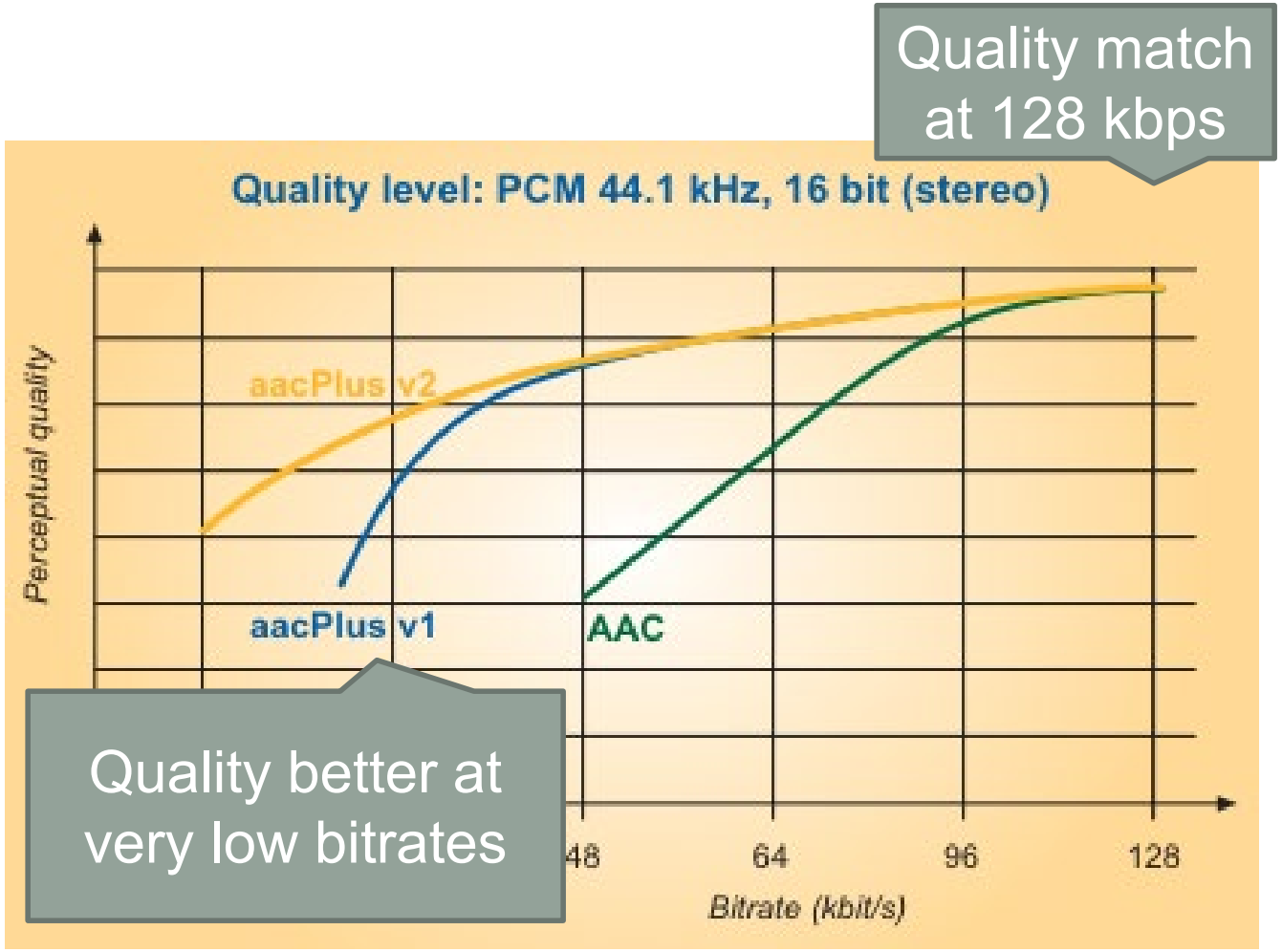
MPEG-4 Audio



- AAC-Low Complexity (AAC-LC)
  - The most basic and most broadly compatible
  - In my tests, indistinguishable from HE AAC/HE AACv2
- High Efficiency AAC (2003)
  - Also called AAC+ and aacPlus
- High Efficiency AACv2 (2006)
  - Also called enhanced AAC+, aacPlus v2 and eAAC+

MPEG-4 Audio Summary

- Performance comparisons
  - aacPlus and aacPlus v2 are low bitrate codecs
  - Superior quality at 96 kbps and below
- Recommendations
  - If 128 kbps stereo (or 64 kbps mono), stay with AAC LC



H.264 Profiles, Levels, and Entropy Coding

- What profiles are and why they exist
  - Quality-related aspects
  - Compatibility aspects
- What levels are
- About entropy coding

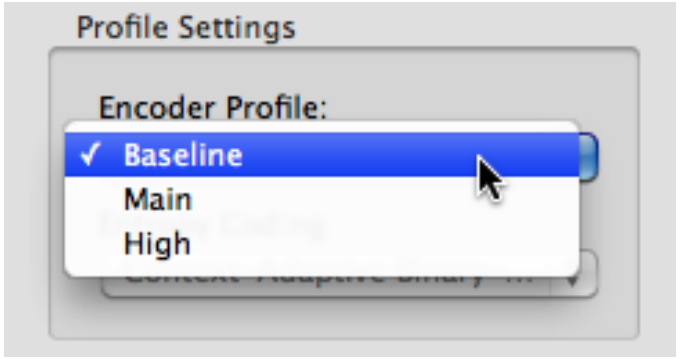
What Profiles are and Why They Exist

- Profiles enable different encoding techniques to balance decoding complexity
- Baseline uses the fewest, so is easiest to decode
  - Early video-capable iPods only supported the Baseline codec
- High uses the most, so is the hardest to decode
  - All computers, mobile devices, TVs, STBs manufactured in the last 6+ years can play the High profile

	Baseline	Main	High
I and P Slices	Yes	Yes	Yes
B Slices	No	Yes	Yes
Multiple Reference Frames	Yes	Yes	Yes
In-Loop Deblocking Filter	Yes	Yes	Yes
CAVLC Entropy Coding	Yes	Yes	Yes
CABAC Entropy Coding	No	Yes	Yes
Interlaced Coding (PicAFF, MBAFF)	No	Yes	Yes
8x8 vs. 4x4 Transform Adaptivity	No	No	Yes
Quantization Scaling Matrices	No	No	Yes
Separate Cb and Cr QP control	No	No	Yes
Separate Color Plane Coding	No	No	No
Predictive Lossless Coding	No	No	No
	Baseline	Main	High

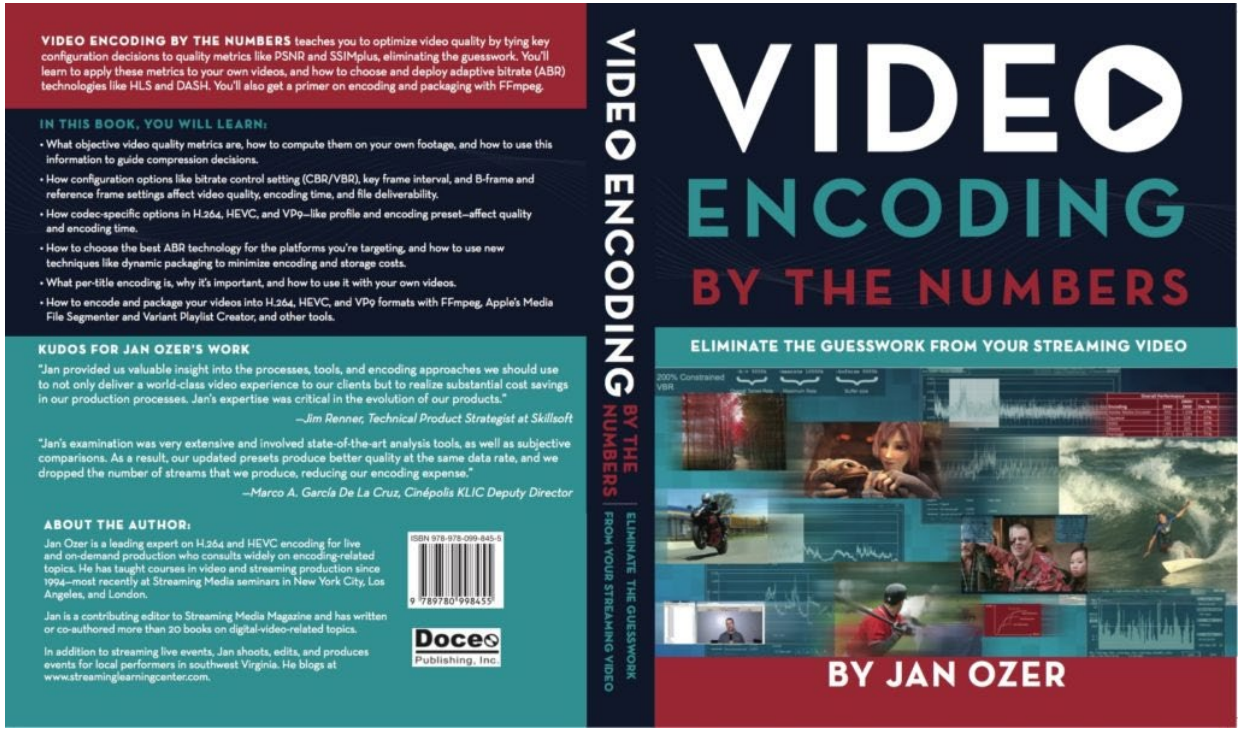
Encoding

- Profiles/Levels
  - Most critical ***compatibility-related*** setting
    - Encode using wrong profile, file won't play on target device
  - Profile is available on all encoding tools
- Don't exceed profile of target device
  - Exclusively a concern with older mobile
  - Computers, OTT devices, and Smart TVs can play High profile (any level)
- What about quality?



Meet Your Test Clips

- Ten files
  - 1 movie (Tears of Steal)
  - 2 animations (Sintel, BBB)
  - Two general purpose (concert, advertisement)
  - One talking head
  - Screencam
  - Tutorial (PPT/Video)
  - New: Football/Soccer
- Why? Because all files are unique and encode differently



- Used for my book Video Encoding by the Numbers
  - http://bit.ly/numbers_book
  - Most of analysis also in book

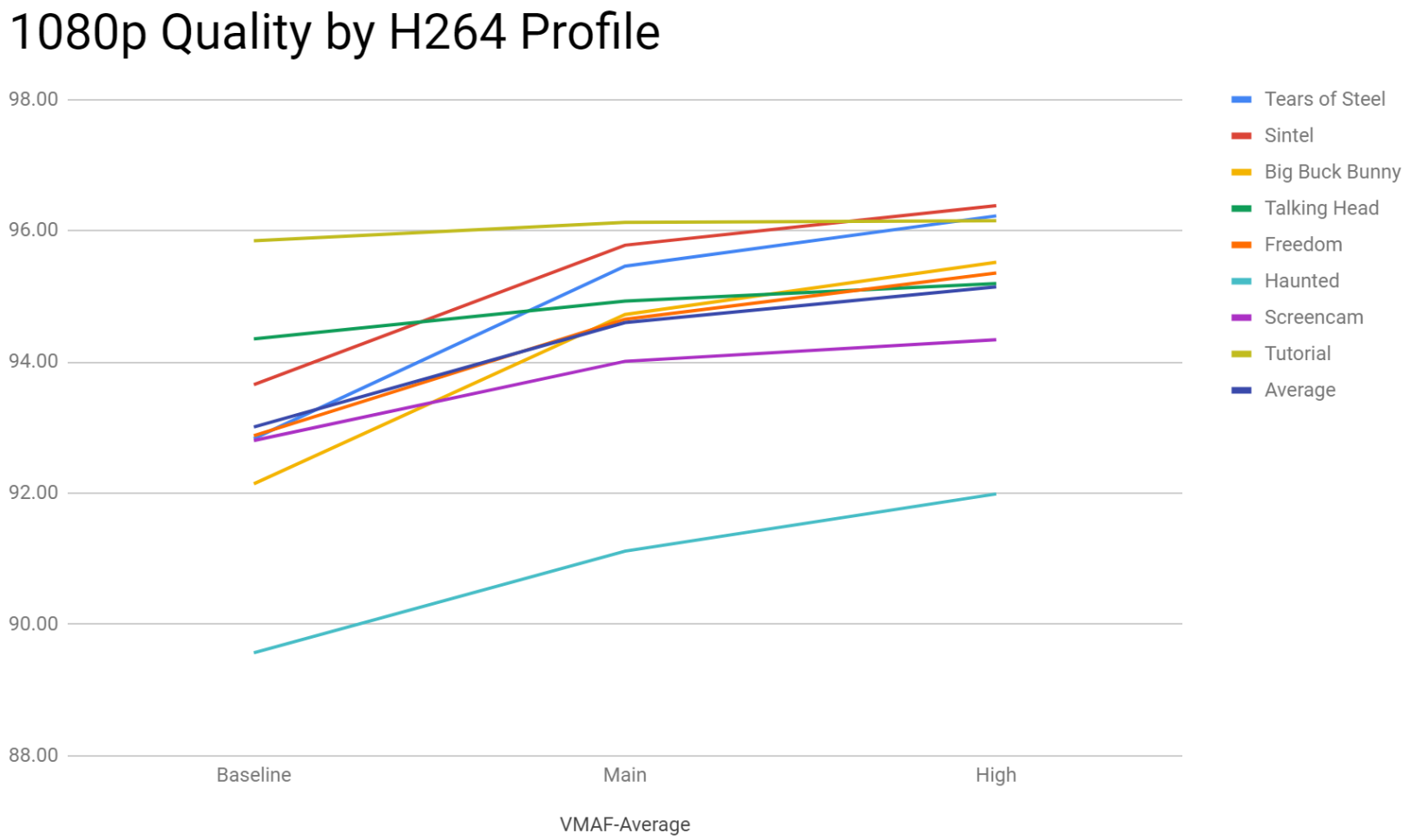
Profiles and Quality

VMAF-Average	Baseline	Main	High	Delta - Baseline/Main	Delta - Main/High	Total Delta
Tears of Steel	92.83	95.46	96.23	2.83%	0.80%	3.66%
Sintel	93.65	95.78	96.38	2.27%	0.63%	2.91%
Big Buck Bunny	92.14	94.72	95.52	2.80%	0.83%	3.67%
Talking Head	94.35	94.93	95.19	0.61%	0.28%	0.90%
Freedom	92.87	94.65	95.36	1.91%	0.74%	2.67%
Haunted	89.56	91.11	91.99	1.73%	0.95%	2.70%
Screencam	92.80	94.01	94.34	1.30%	0.35%	1.66%
Tutorial	95.85	96.13	96.15	0.29%	0.03%	0.32%
Average	**93.01**	**94.60**	**95.14**	**1.72%**	**0.57%**	**2.31%**

- High is always the best; Baseline always the worst
  - Jump from Baseline > Main more significant than Main > High
- Difference is greater in hard to encode files
  - TOS – 3.66%
  - Talking Head – .9%

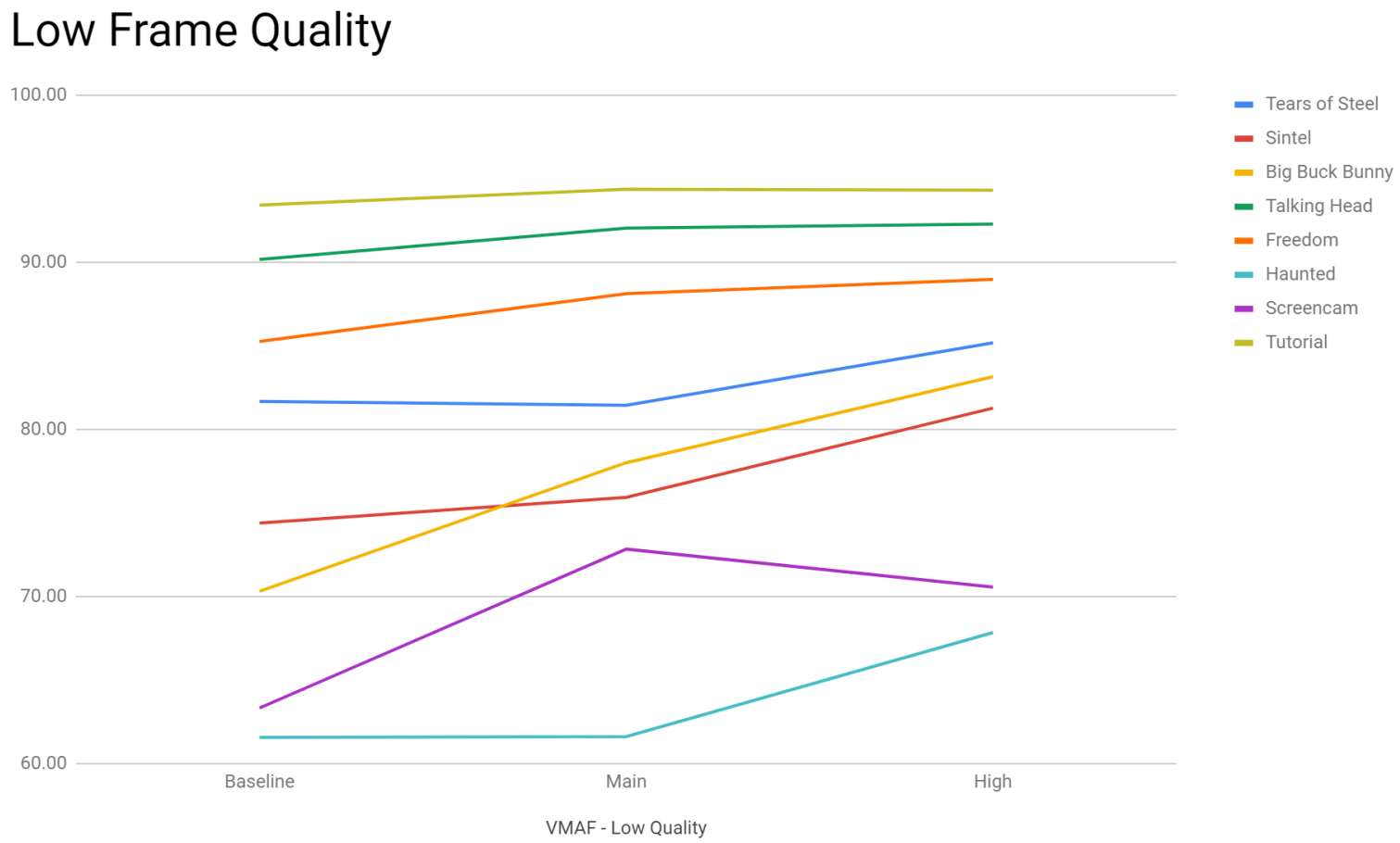
Profiles and Quality

- High is always the best; Baseline always the worst
  - Jump from Baseline > Main more significant than Main > High
- Difference is greater in hard to encode files
  - TOS – 3.66%
  - Talking Head – .9%



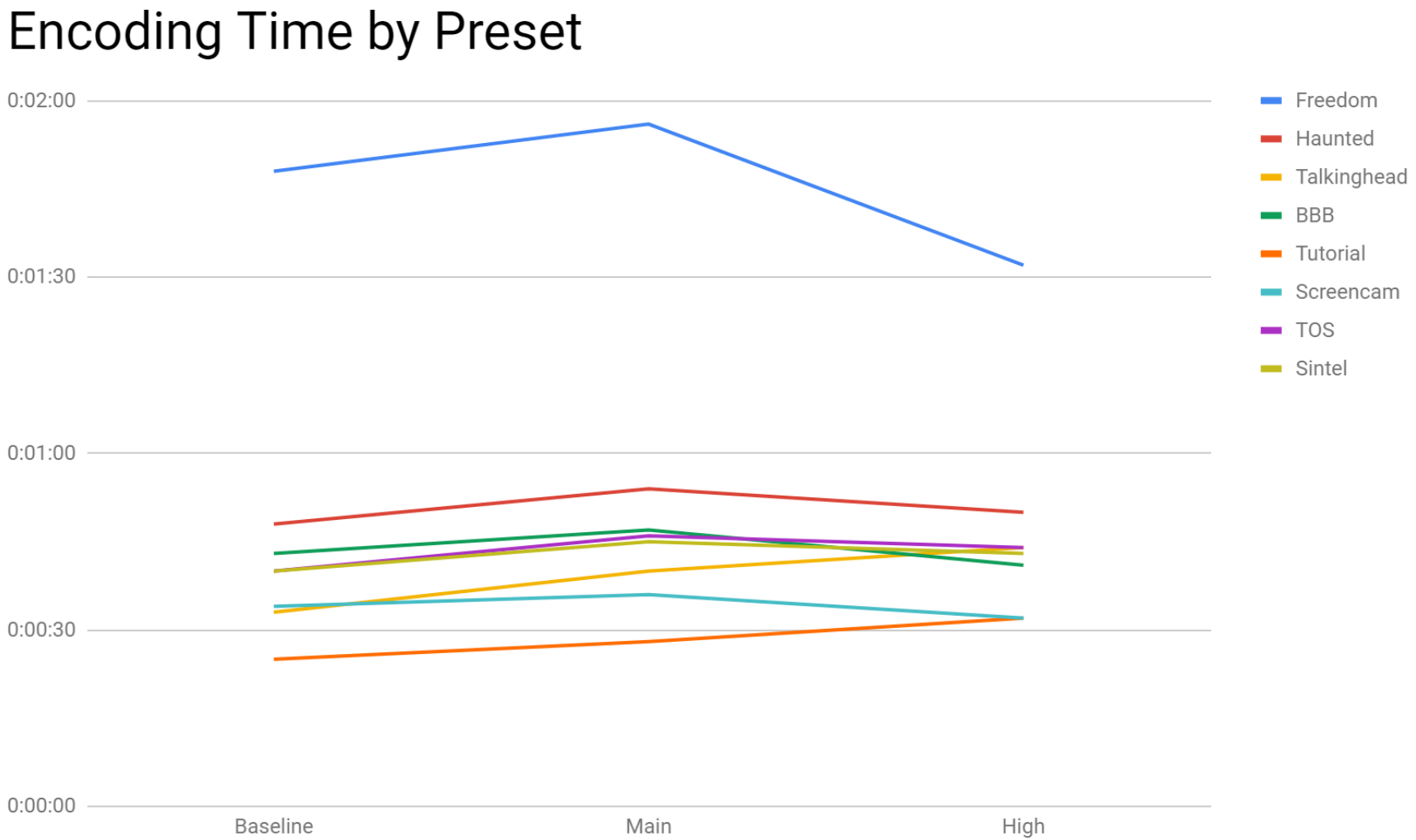
Profiles and Low Frame Quality

- Low frame measures potential for transient quality issues
- VMAF score of lowest quality frame in the file
  - Baseline is always the worst; sometimes significantly so
- Max deltas
  - Big Buck Bunny – 12.85 (2x JND)
  - Sintel – 6.88
  - Haunted (DSL production) – 6.28



Profiles and Encoding Time

- Baseline is fastest, but High is faster than Main
- Cheaper to encode to High than Main (though difference is minimal)



How to Use Profiles in Encoding Ladder

- History lesson
- Quality deltas
- Decision tree

iOS History Lesson

Baseline only devices are really old (really old)

Width	Height	Frame Rate	Video Bitrate	Audio Bitrate	I-Frame	Profile	B-frames	Segment Size	iPod Touch 2-4	iPod Touch 5	iPhone 3G, 3GS, 4	iPhone 4S, 5, 5C, 5S	iPad 1,2	iPad 3, 4, 5	Apple TV 2	Apple TV 3
416	234	12	200	64	36	Baseline	NA	9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
480	270	15	400	64	45	Baseline	NA	9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
640	360	29.97	600	64	90	Baseline	NA	9	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
640	360	29.97	1200	96	90	Baseline	NA	9		Yes		Yes	Yes	Yes	Yes	Yes
960	540	29.97	3500	96	90	Main		9		Yes		Yes	Yes	Yes	Yes	Yes
1280	720	29.97	5000	128	90	Main	As	9		Yes		Yes	Yes	Yes	Yes	Yes
1280	720	29.97	6500	128	90	Main	needed	9		Yes		Yes	Yes	Yes	Yes	Yes
1920	1080	29.97	8500	128	90	High		9		Yes		Yes		Yes		Yes

- Initial version of TN2224 customized profile for different targets

Current HLS Authoring Specs Abandon Legacy Devices

HDR (HEVC) 30 fps	HEVC/H.265 30 fps	H.264/AVC	Resolution 16:9 aspect ratio	Frame rate
160	145	145	416 x 234	≤ 30 fps
360	300	365	480 x 270	≤ 30 fps
800	660	730	640 x 360	≤ 30 fps
1200	990	1100	768 x 432	≤ 30 fps
2050	1700	2000	960 x 540	same as source
2900	2400	3000	1280 x 720	same as source
3850	3200	4500	1280 x 720	same as source
5400	4500	6000	1920 x 1080	same as source
7000	5800	7800	1920 x 1080	same as source
9700	8100	n/a	2560 x 1440	same as source
13900	11600	n/a	3840 x 2160	same as source
20000	16800	n/a	3840 x 2160	same as source

- Significant change in 2017:
  - Expect all to play High profile

http://bit.ly/A_Devices_Spec

Encoding for Android Devices

The table below lists the Android media framework video encoding profiles and parameters recommended for playback using the H.264 Baseline Profile codec. The same recommendations apply to the Main Profile codec, which is only available in Android 6.0 and later.

	SD (Low quality)	SD (High quality)	HD 720p (N/A on all devices)
Video resolution	176 x 144 px	480 x 360 px	1280 x 720 px
Video frame rate	12 fps	30 fps	30 fps
Video bitrate	56 Kbps	500 Kbps	2 Mbps
Audio codec	AAC-LC	AAC-LC	AAC-LC
Audio channels	1 (mono)	2 (stereo)	2 (stereo)
Audio bitrate	24 Kbps	128 Kbps	192 Kbps

- Android support is bifurcated
  - In OS software – Baseline profile only
  - In hardware/device supplied software, up to High
- Google recommends using Baseline or Main (bit.ly/androidvideospecs)
  - Ignored by many

Quality Difference in **b** - Talking Head Clip

Talking Head - VMAF	Baseline	Main	High	Baseline /High	Baseline /Main	Main/ High
1080p	94.09	94.80	95.08	-0.99	-0.71	-0.28
720p	89.85	90.41	90.52	-0.68	-0.56	-0.11
540p	85.90	86.40	86.48	-0.58	-0.50	-0.09
480p	83.32	84.03	84.00	-0.68	-0.71	0.03
360p	74.77	75.83	75.79	-1.03	-1.06	0.04
270p	55.73	57.42	57.46	-1.73	-1.70	-0.04
180p	22.54	24.48	24.53	-1.99	-1.95	-0.04
Average	72.31	73.34	73.41	**-1.10**	**-1.03**	**-0.07**

Most viewers watch these rungs

Main/High delta is tiny

Baseline the worst

High the best

Delta greater in lowest rungs

Synthesis:

- Baseline in lower rungs penalizes viewers on newer devices
  - Usually very small % of viewers
- Main/High delta irrelevant

Quality Difference in **Encoding Ladder** – Football Clip

Delta greater in highest rungs

Football - VMAF	Baseline	Main	High	Baseline /High	Baseline /Main	Main/ High
1080p	91.22	94.68	95.59	-4.36	-3.46	-0.90
720p	66.33	73.24	75.10	-8.77	-6.91	-1.86
540p	59.62	66.26	67.69	-8.07	-6.64	-1.43
480p	49.47	55.76	57.17	-7.70	-6.29	-1.41
360p	38.09	43.10	44.01	-5.92	-5.01	-0.91
270p	24.56	27.76	28.12	-3.56	-3.20	-0.36
180p	8.81	10.51	10.27	-1.47	-1.70	0.24
Average	48.30	53.05	53.99	**-5.69**	**-4.75**	**-0.95**

Most viewers watch these rungs

Baseline the worst

High the best

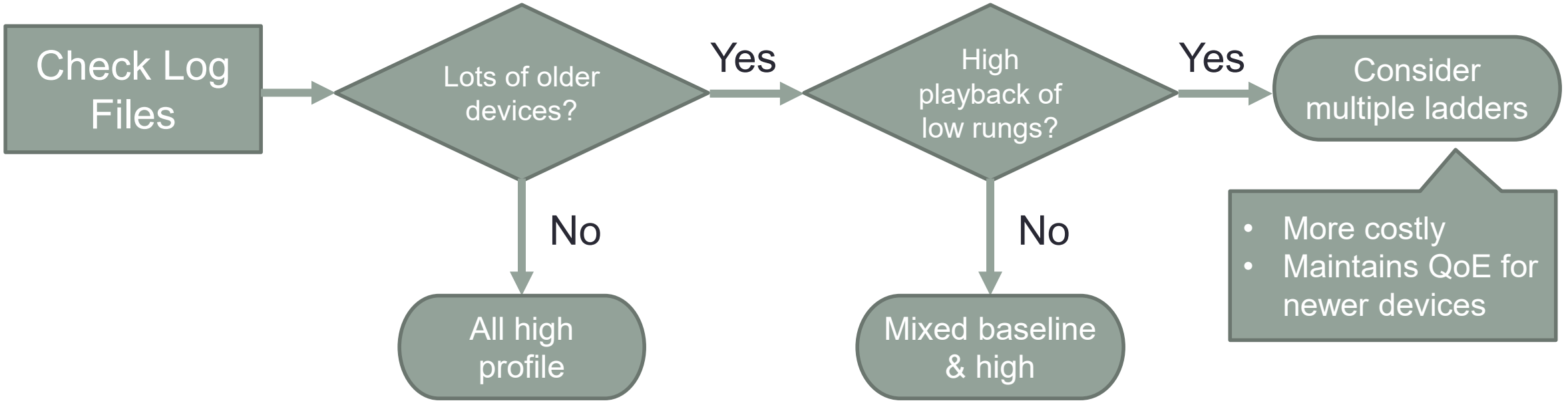
Lower rung deltas significant

Main/High delta still minimal

Synthesis:

- Baseline in lower rungs penalizes viewers on newer devices
  - Usually very small % of viewers
- Need High in top ladder rungs

Encoding for Legacy Devices – Decision Flow



	Profile
1080p	High
720p	High
540p	High
480p	High
360p	High
270p	Baseline
180p	Baseline

- Cheaper
- Minimal QoE loss
- Almost all devices that can play main can play high as well

Conclusions

- More and more, publishers DON’T customize streams for different targets; either:
  - Go High profile and abandon legacy (really iPhone 3 and previous)
  - Use one set of streams with mixed profiles
- Justification
  - Better to retain viewers on legacy devices, even if lower rung quality suffers somewhat
  - Lower rungs aren’t accessed all that much anyway

What Levels are and Why They Exist

Video formats supported: H.264 video up to 4K, 30 frames per second, High Profile level 4.2 with AAC-LC audio up to 160 Kbps, 48kHz, stereo audio or Dolby Audio up to 1008 Kbps,

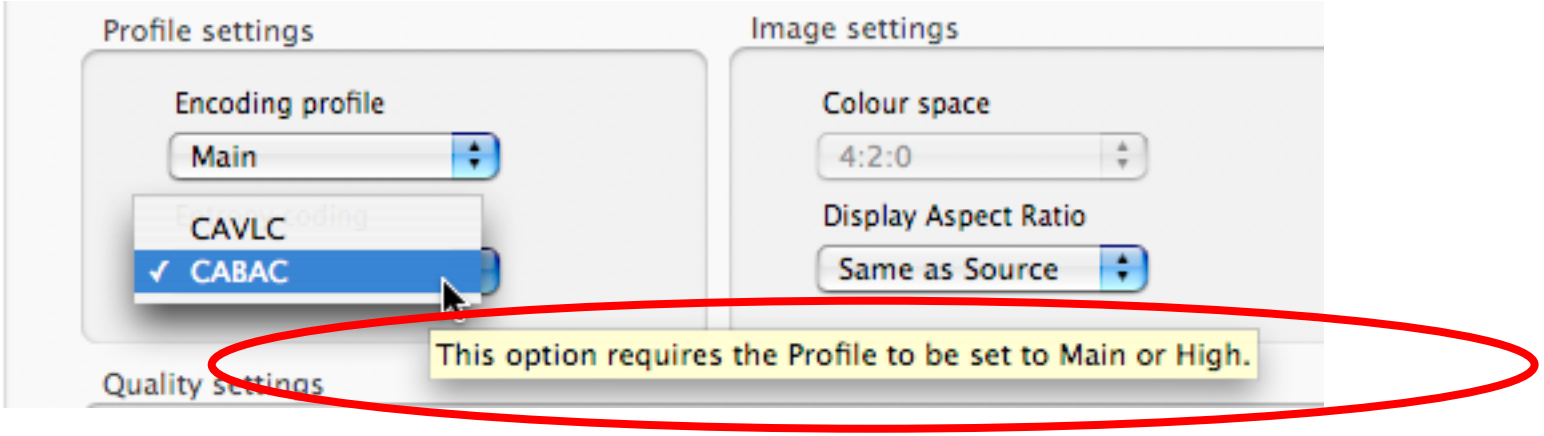
- Levels set further limits on how video can be produced
  - See above specs for the iPhone 7
  - If your encoded videos exceed level spe
- In general, solely a concern for legacy mobile devices

Levels from Wikipedia

Level	Max decoding speed		Max frame size		Max video bit rate for video coding layer (VCL) kbit/s			Examples for high resolution @ highest frame rate (max stored frames)
	Luma samples/s	Macroblocks/s	Luma samples	Macroblocks	Baseline, Extended and Main Profiles	High Profile	High 10 Profile	Toggle additional details
4	62,914,560	245,760	2,097,152	8,192	20,000	25,000	60,000	2,048×1,024@30.0 (4
4.1	62,914,560	245,760	2,097,152	8,192	50,000	62,500	150,000	2,048×1,024@30.0 (4
4.2	133,693,440	522,240	2,228,224	8,704	50,000	62,500	150,000	2,048×1,080@60.0 (4

- You must make sure to keep your encodes within these constraints
  - Otherwise, video might not play
- Again, typically only a concern for very old legacy devices

Entropy Encoding



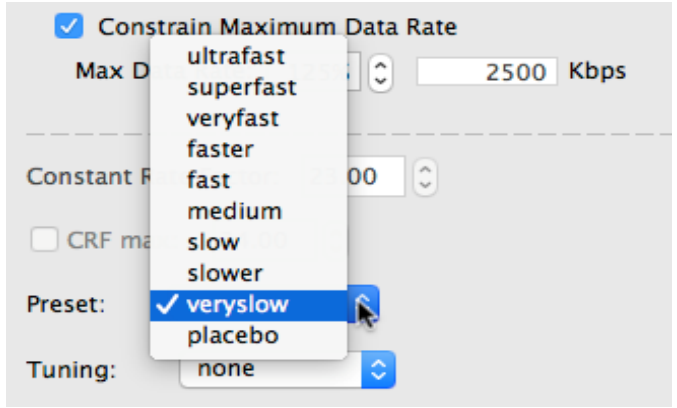
- How the compressed data is packed
  - CAVLC for Baseline (only option)
  - CABAC for Main/High profiles (highest quality option)

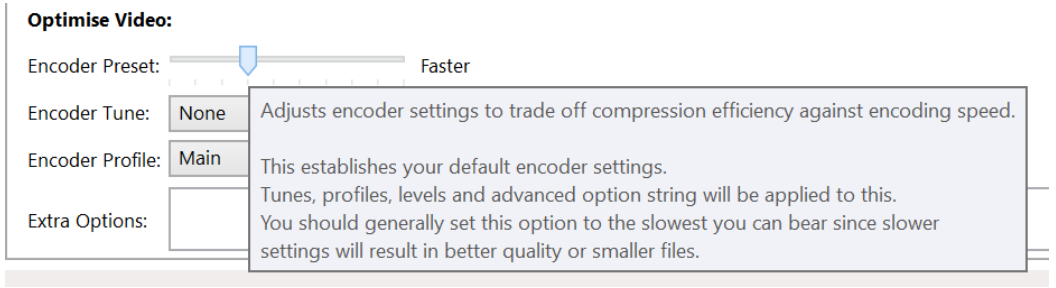
Lesson: Choosing an x264 Preset

- What's a preset?
- How to choose the best one
  - Quality
  - Time
- Application

Choosing an x264 Preset

- What are presets?
  - Simple way to adjust multiple parameters to balance quality and encoding time
  - This analysis – x264 presets
  - Most H.264 codecs have something similar
- Medium is generally the default x264 preset
  - Is this the best for you?
- Relevant for encoders using x264 codec
  - Other encoders may have a similar quality/encoding time toggle should run similar analysis





Average Quality

Average Quality	Ultrafast	Superfast	Veryfast	Faster	Fast	Medium	Slow	Slower	Veryslow	Placebo	Total Delta
Tears of Steel	89.20	92.00	93.29	95.45	95.59	96.22	96.43	96.56	96.67	96.65	8.38%
Sintel	88.29	92.66	93.85	95.84	95.99	96.38	96.56	96.68	96.83	96.75	9.68%
Big Buck Bunny	87.26	91.26	92.68	95.03	95.29	95.53	95.75	95.87	96.05	96.01	10.08%
Talking Head	95.19	92.55	93.66	94.90	94.86	95.18	95.29	95.43	95.51	95.39	3.20%
Freedom	91.95	91.15	92.63	94.58	94.51	95.37	95.59	95.84	96.15	96.04	5.48%
Haunted	91.30	88.61	89.43	91.30	91.08	91.98	92.08	92.35	92.49	92.45	4.38%
Screencam	90.92	92.56	93.52	94.75	94.75	94.70	94.77	94.86	94.92	94.91	4.41%
Tutorial	93.42	94.66	95.55	96.16	96.17	96.17	96.26	96.28	96.29	96.10	3.07%
Average	**90.53**	**91.37**	**92.59**	**94.52**	**94.56**	**95.11**	**95.28**	**95.46**	**95.62**	**95.55**	**6.08%**

- Red is lowest quality
- Green highest quality
- Note top values – average 95.62 (not Placebo)
- Very slow averages best quality
  - But only 6% spread between best and worst

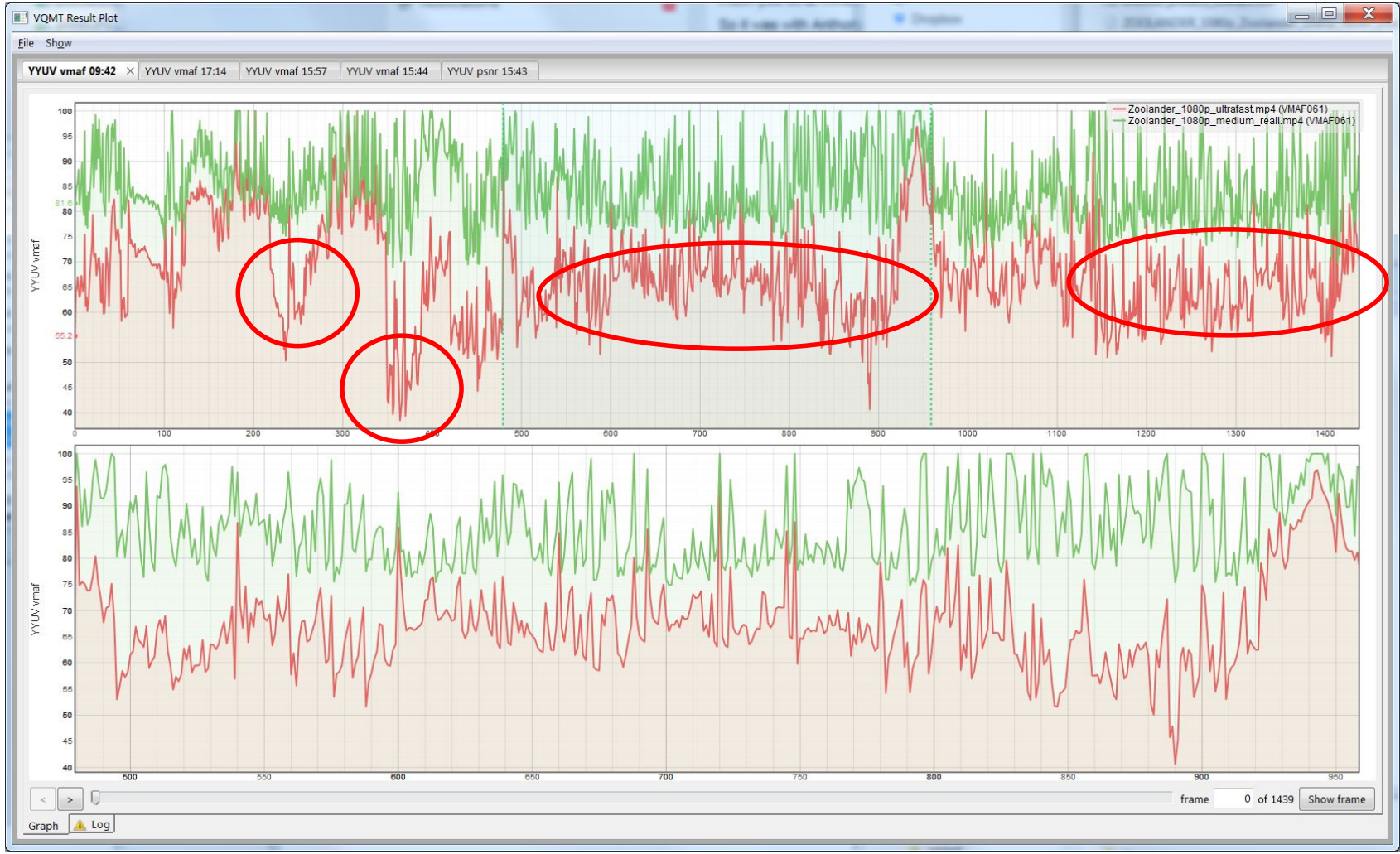
Low Frame Quality

Low Frame Quality	Ultrafast	Superfast	Veryfast	Faster	Fast	Medium	Slow	Slower	Veryslow	Placebo	Total Delta
Tears of Steel	70.16	74.82	77.67	84.51	85.02	85.34	85.44	86.38	85.33	85.10	23.12%
Sintel	68.77	69.79	74.93	79.12	80.41	82.27	81.90	82.98	84.89	82.61	23.45%
Big Buck Bunny	55.42	65.11	62.50	79.33	79.57	82.70	79.18	83.22	80.24	79.08	50.15%
Talking Head	88.90	61.43	88.53	91.62	91.32	92.11	92.03	92.49	92.16	91.37	50.56%
Freedom	76.49	82.79	83.96	87.59	87.29	88.72	89.00	89.35	90.28	90.05	18.03%
Haunted	60.36	57.18	62.69	64.62	61.63	67.33	67.74	68.64	72.08	72.28	26.42%
Screencam	56.16	68.53	71.00	76.39	77.44	77.06	78.04	79.26	78.04	75.21	41.12%
Tutorial	85.68	90.99	91.95	94.11	94.24	94.68	94.50	94.21	94.02	70.58	34.15%
Average	**70.02**	**68.52**	**75.05**	**81.13**	**80.88**	**83.08**	**82.55**	**83.84**	**84.16**	**83.41**	**33.37%**

- Red is lowest quality
- Green highest quality
- Very slow averages best quality
  - **33.37% difference from best to worst**

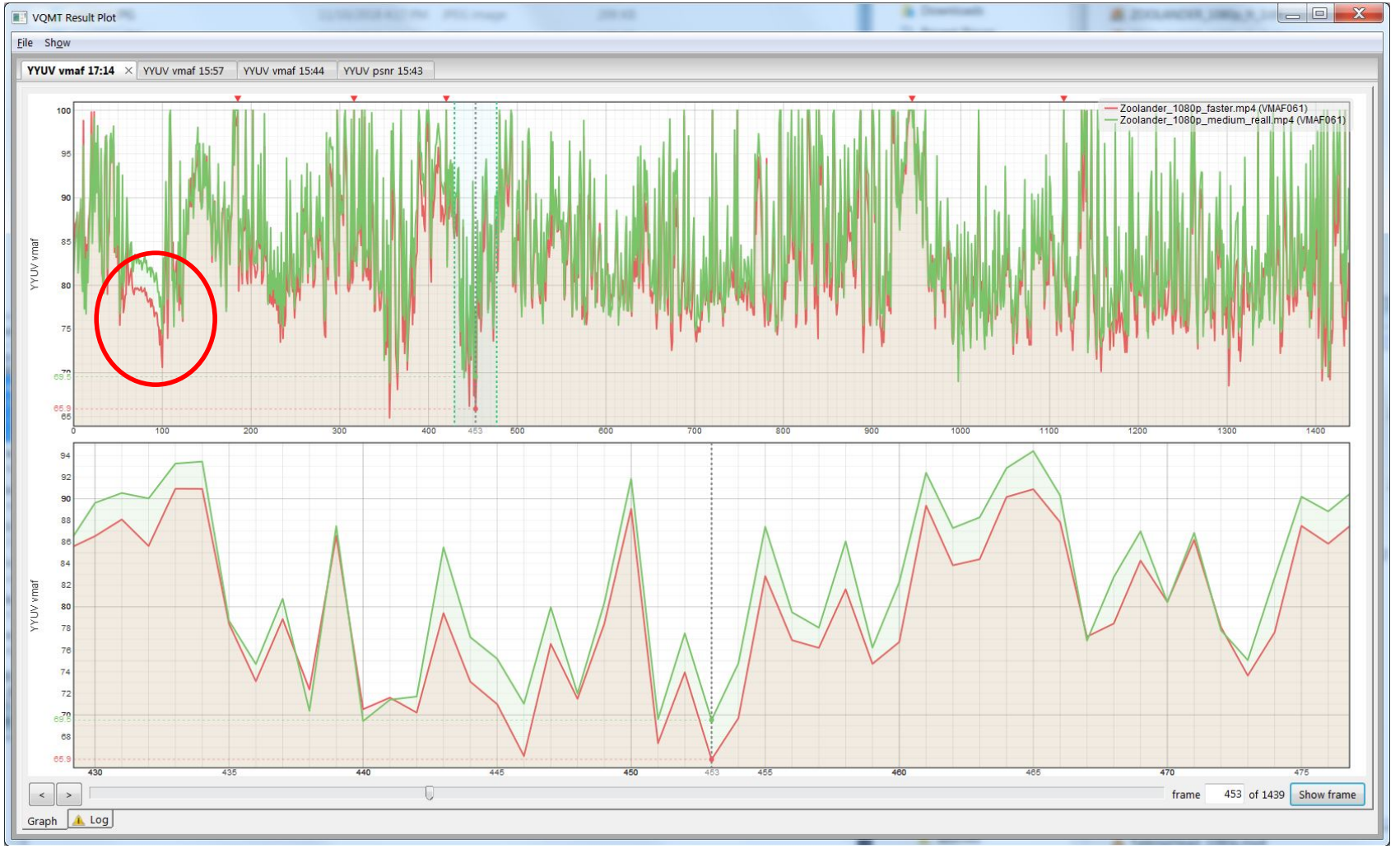
Check Results Plot – Ultrafast (red) vs Medium

- Plot of VMAF values over duration of clip
  - Red is ultrafast
  - Green is Medium
- Multiple deep drops that would be noticeableNever use ultrafast (even in live)

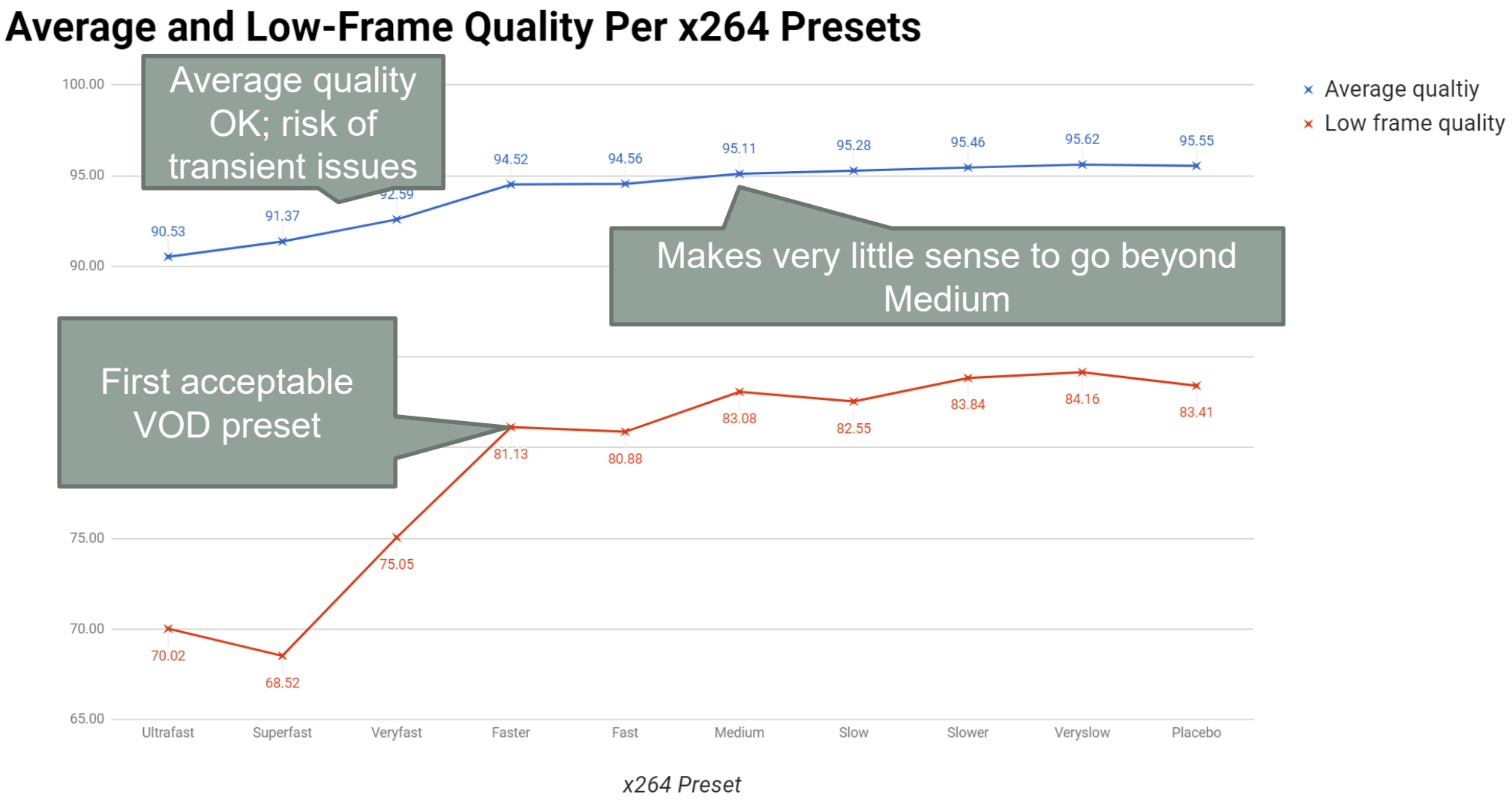


Check Results Plot – Faster (red) vs Medium

- One problem area, but no major quality differences
- Faster should be acceptable starting point for VOD and live
  - Cut encoding time by over 66% with no quality hit
  - Said another way, triple capacity



Average and Low Frame Graphed

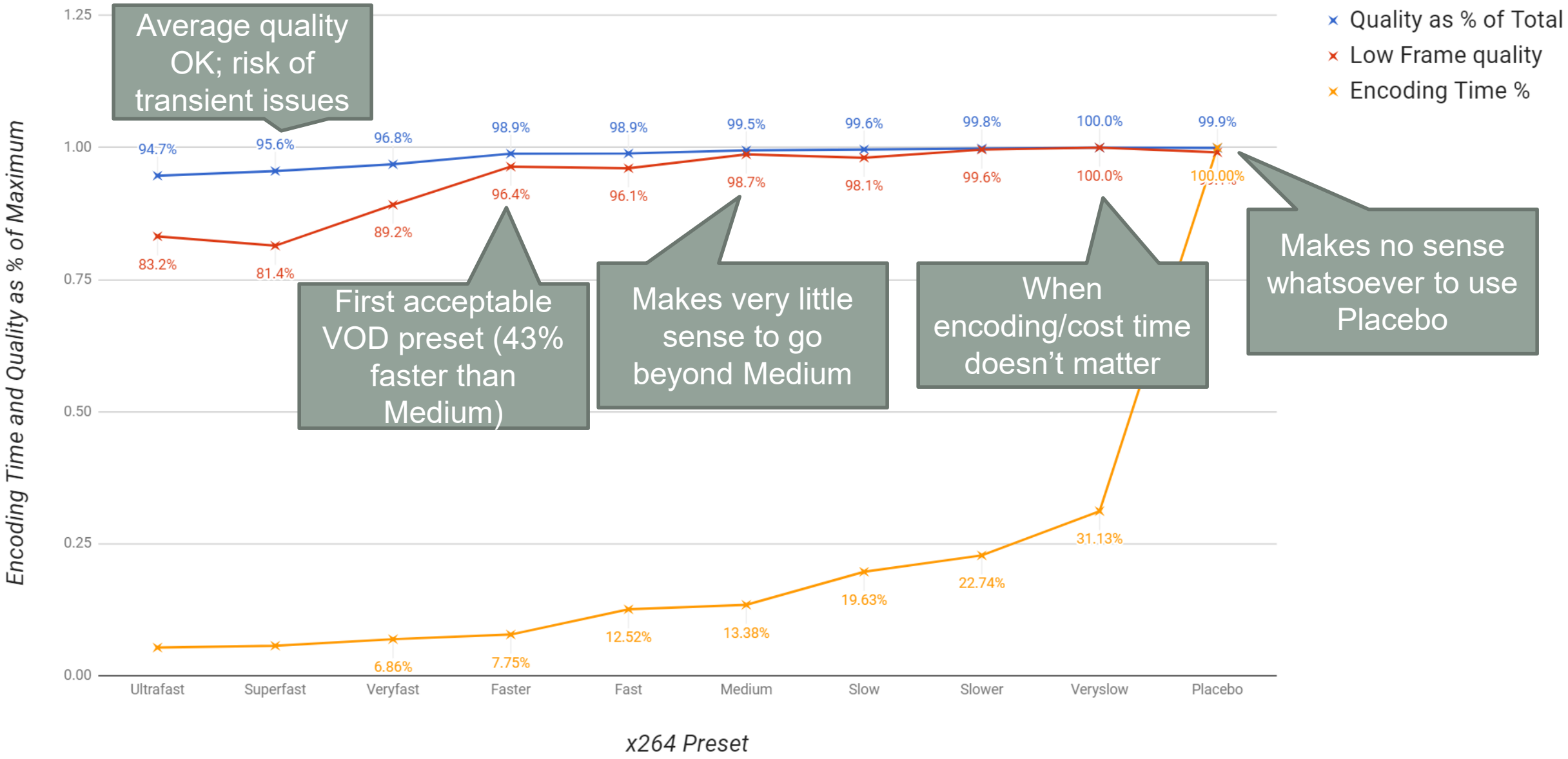


Encoding Time

	Ultrafast	Superfast	Veryfast	Faster	Fast	Medium	Slow	Slower	Veryslow	Placebo
Tears of Steel	9	10	14	16	28	30	47	57	72	226
Sintel	9	10	13	16	28	30	46	57	72	241
Big Buck Bunny	12	13	14	16	29	31	49	57	76	288
Talking Head	11	12	14	16	27	30	40	51	61	237
Freedom	11	12	16	20	36	40	59	45	90	273
Haunted	14	15	17	22	37	40	60	73	92	288
Screencam	11	12	15	13	15	14	18	29	43	79
Tutorial	13	12	14	13	13	13	15	18	24	71
Average	**11**	**12**	**15**	**16**	**27**	**28**	**42**	**48**	**66**	**213**
Percentage of slowest	**5.28%**	**5.62%**	**6.86%**	**7.75%**	**12.52%**	**13.38%**	**19.63%**	**22.74%**	**31.13%**	**100.00%**

- Seconds per encode
- Percentage of slowest
  - Ultrafast was 5.28% of Placebo
  - Medium was 13.38%

Average Quality, Low-Frame Quality and Encoding Time Per x264 Presets



Conclusions

- Faster is best preset for those seeking maximum throughput
- Makes very little sense to go beyond Medium when encoding cost/time is a concern
- Veryslow delivers maximum average and low-frame quality
- Placebo never seems to make sense

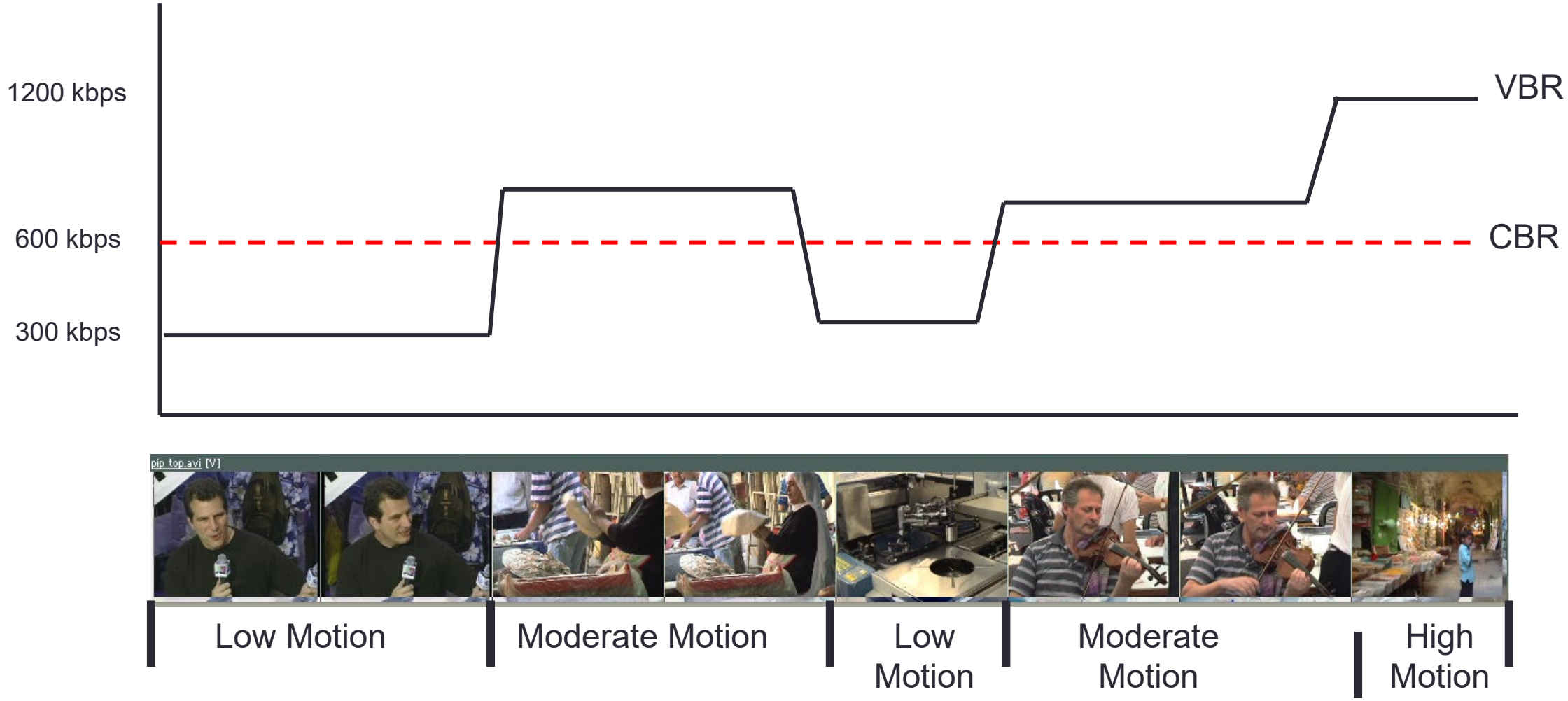
Bottom Line

- For those encoding with x264, medium may not be the best preset if you're reaching encoding capacity
- Those using other encoders:
  - Find the quality/encoding time toggle
  - Run similar tests
  - Find the optimal setting

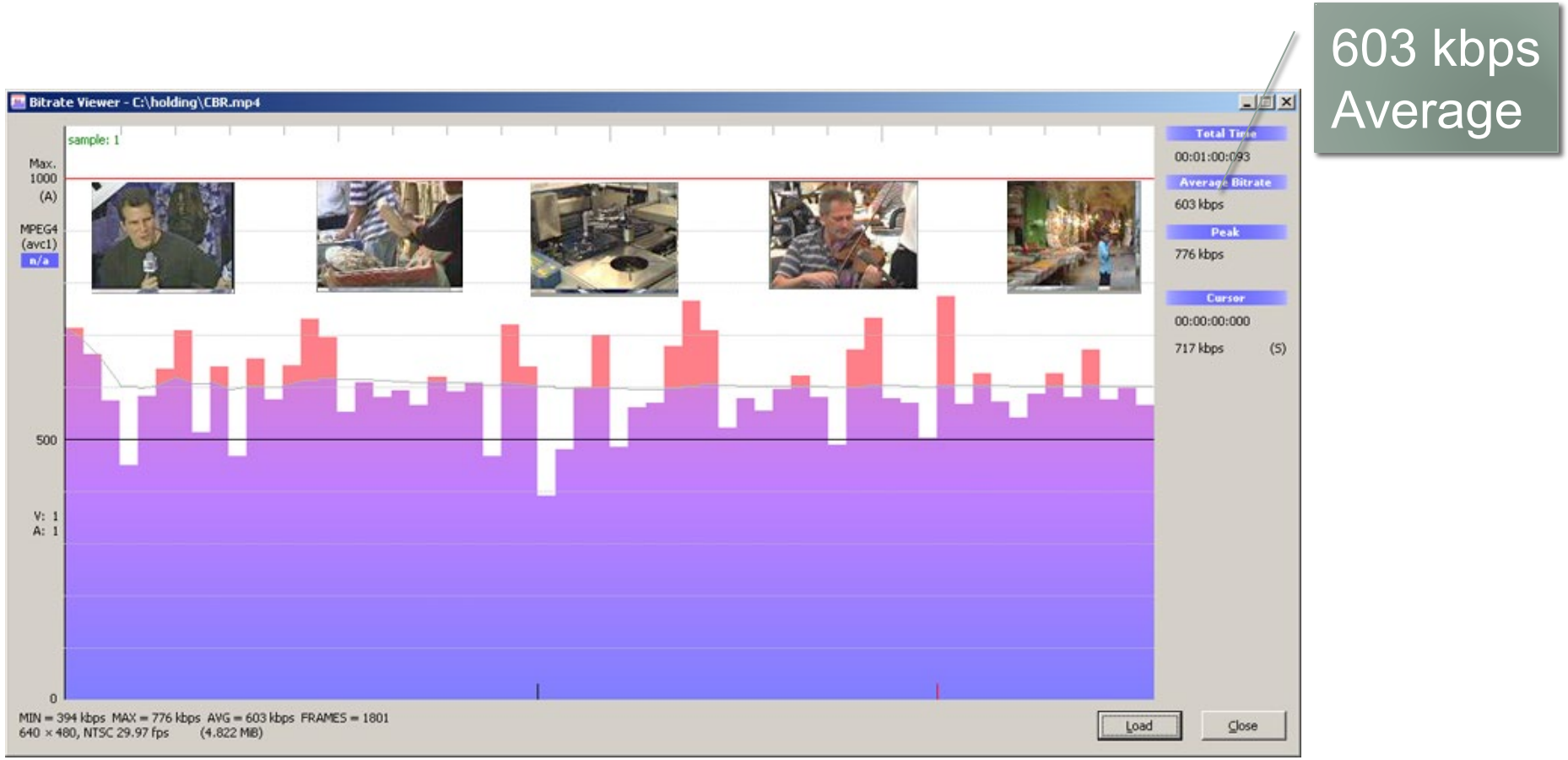
Lesson: Bitrate Control

- How VBR and CBR work
- Differences in overall and transient quality
- How both techniques affect deliverability

How VBR and CBR Work



CBR File Illustrated



- Faint (sorry) wavy blue line is data rate
- Relatively consistent throughout

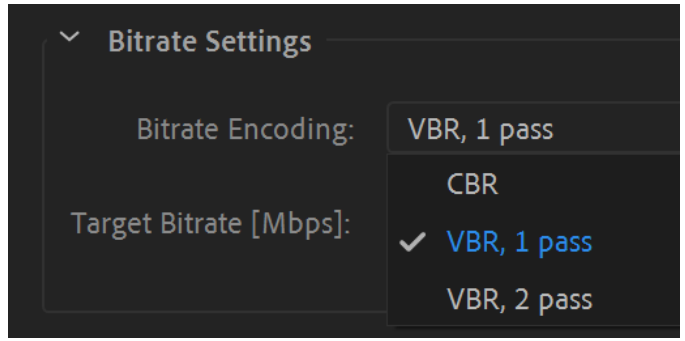
VBR File Illustrated

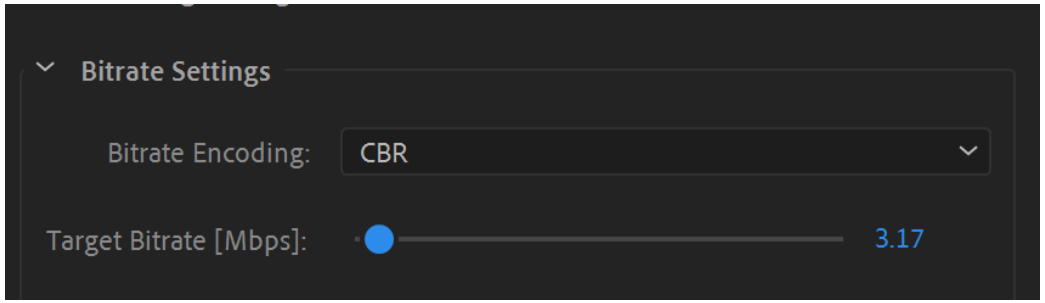
596 kbps Average

- Faint (sorry) wavy blue line is data rate
- Varies with scene complexity

Producing CBR

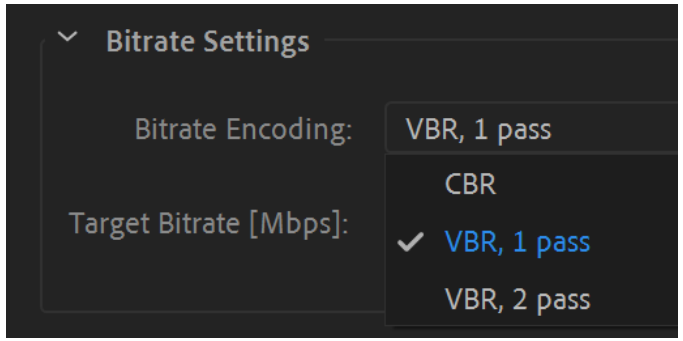
- Typical uses:
  - Live
  - Streaming to constrained lower bitrate connections like 3G
- Typically, single-pass, but can be two-pass
  - Choose CBR, then choose target bitrate

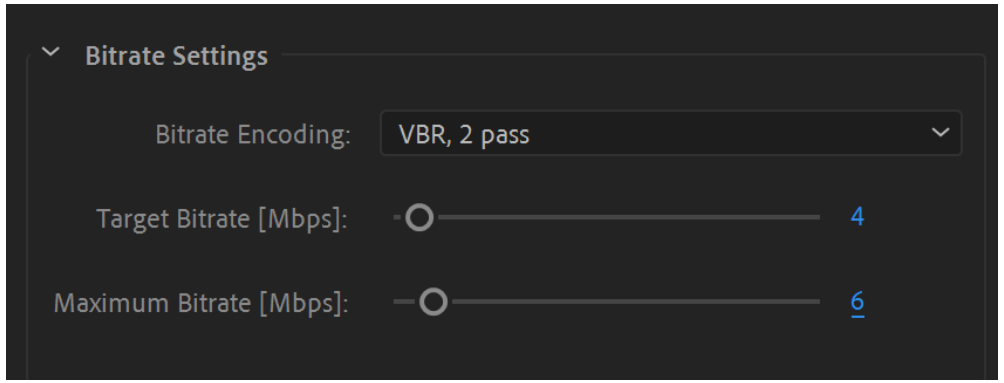




Producing VBR

- Typical uses
  - Most VOD streaming
  - Most mezz file creation
- Typically, two-pass, but can be single or multiple
  - Choose VBR, then choose:
    - Target
    - Maximum (1.1x – 2x, here 1.5x)
    - Sometimes minimum (typically .5x)





How Much Better Quality is VBR over CBR?

VMAF	200% VBR	150% VBR	110% VBR	2-Pass CBR	1-Pass CBR	Total Delta	Delta 110%-200%
Tears of Steel	97.6	97.1	97.2	97.1	97.0	0.53	-0.34
Sintel	97.6	97.5	97.7	97.6	97.2	0.56	0.11
Big Buck Bunny	96.8	96.5	95.9	95.6	95.9	1.18	-0.89
Talking Head	95.7	95.7	95.7	95.7	95.7	0.09	-0.03
Freedom	97.6	96.7	96.6	96.3	96.4	1.32	-1.05
Haunted	94.5	94.5	94.4	94.3	94.5	0.24	-0.10
Screencam	95.5	95.3	94.9	94.3	95.2	1.18	-0.66
Tutorial	97.3	97.2	97.1	97.1	97.1	0.19	-0.12
Average	**96.6**	**96.3**	**96.2**	**96.0**	**96.1**	**0.58**	**-0.38**

- Across the spectrum of content – overall score - not that much – average .58 VMAF at 1080p

What About Transient Issues - Look at Results Plot

- Red is first file (CBR)
- Green is second (200% constrained VBR)
- Graph tracks ratings over entire file
- Top graph is entire file
- Bottom graph is expanded view of selected region up top
- Circled areas shows very significant quality delta

Let’s Look at Frames - Original

Let’s Look at Frames – Constrained VBR

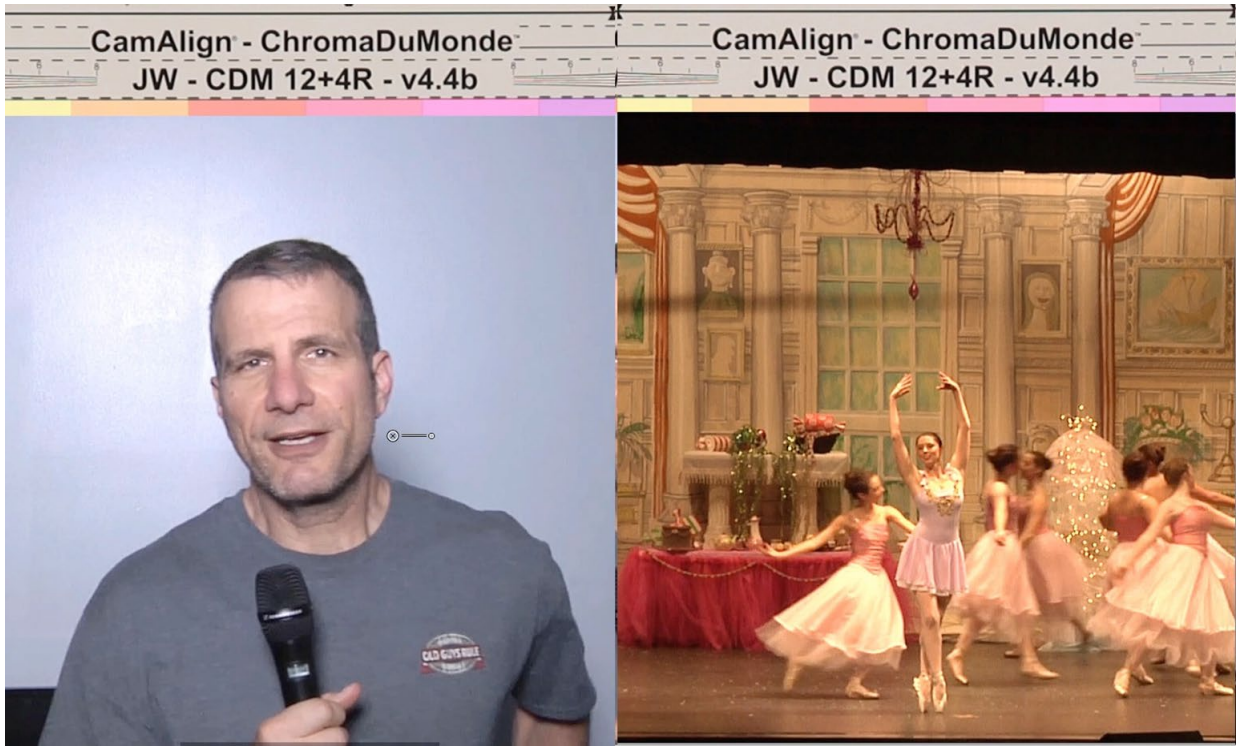
Let’s Look at Frames - CBR

Bottom Line - Quality

- Overall quality difference is minor
- CBR has potential for more transient quality issues
- Constrained VBR – 110% - 200% typically avoids these

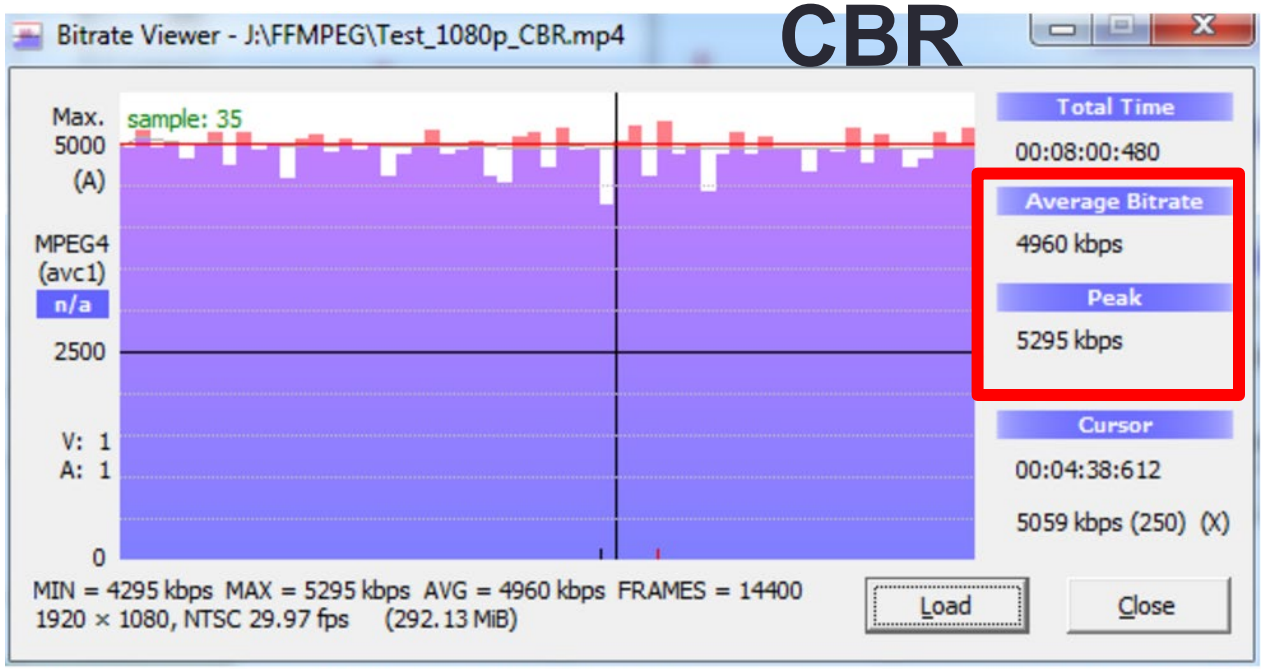
What About Deliverability

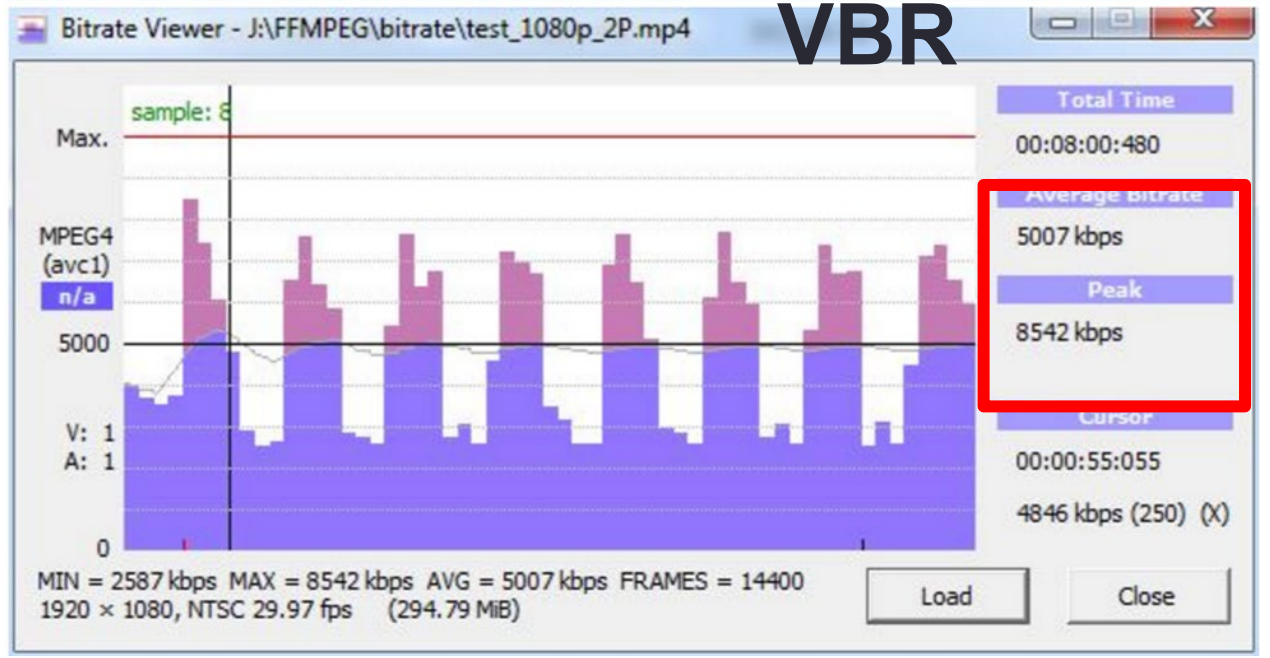
- VBR can create files that may be difficult to deliver over fixed connections
- This worst-case test file combines 30-seconds of talking head with 30-seconds of ballet



Deliverability

- Which file is easier to deliver over fixed bandwidth connections?
  - Overall bitrate very similar (CBR slightly higher)
  - But data rate is much more predictable, and therefore easier to deliver
- So, limit variability by implementing *constrained* VBR
  - Limit peaks to % over target
  - Most practitioners limit to 200%





CBR/VBR Summaries

Constant Bitrate

- Pros:
  - Easiest stream to deliver
- Cons
  - Lowest overall quality
  - Transient quality issues
- Best application
  - Live streaming

Variable Bitrate

- Pros:
  - Best overall quality
  - No transient quality problems
- Cons
  - Can cause deliverability issues
- Best application
  - VOD

Constant Rate Factor (CRF) and Capped CRF

- About CRF encoding
- Encoding with CRF
- Capped CRF

On Constant Rate Factor Encoding

- What is it:
  - An encoding mode in x264, x265, VP9
  - Adjusts data rate to achieve target quality
  - Quality range is 1-51; lower levels are higher quality

```
FFmpeg –i input.mp4  -b:v 5000k output.mp4
```

Delivers 5 Mbps; quality varies

```
FFmpeg –i input.mp4  -crf 23 output.mp4
```

Delivers crf 23 quality; bitrate varies

Constant Rate Factor (CRF) Encoding

- VBR and CBR encodes attempt to hit the specified data rate
  - Adjusts quality to hit the target data rate
- CRF delivers specified quality (on a scale from 1-51)
  - Adjusts data rate to hit the specified quality level
- You can’t use CRF for streaming (because no data rate limit)
  - Useful for archiving

Capped CRF Encoding

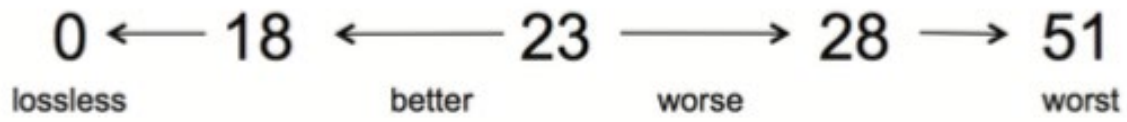
- Capped CRF combines CRF encoding with a data rate cap
  - Delivers specified quality with a maximum bitrate
  - Can be single pass
  - Used by some large OTT shops and OVPs as data rate control mechanism for fast/cheap per-title encoding
- Key benefit
  - Delivers very good quality at low data rates for easy-to-compress videos
  - Customizes encodes depending upon video quality – form of per-title encoding

Producing Capped CRF

```
ffmpeg -i Test_1080p.mp4 -c:v libx264 -crf 23 -maxrate 3500k -bufsize 7000k
Test_CRF23_3500.mp4
```

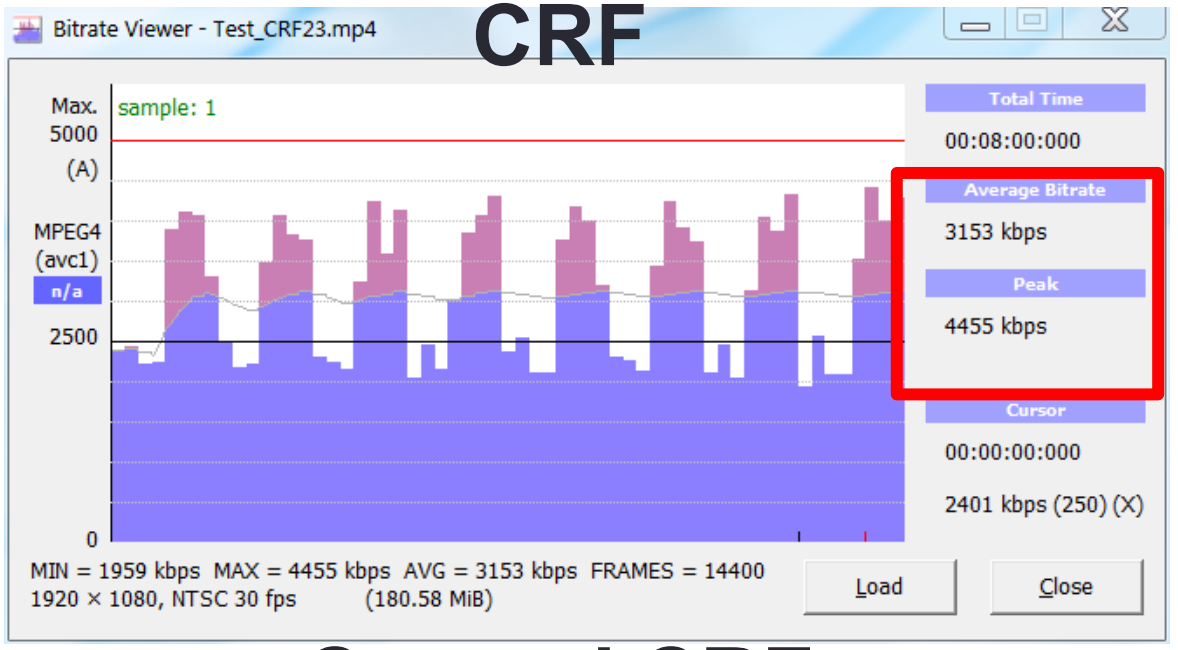
Commands

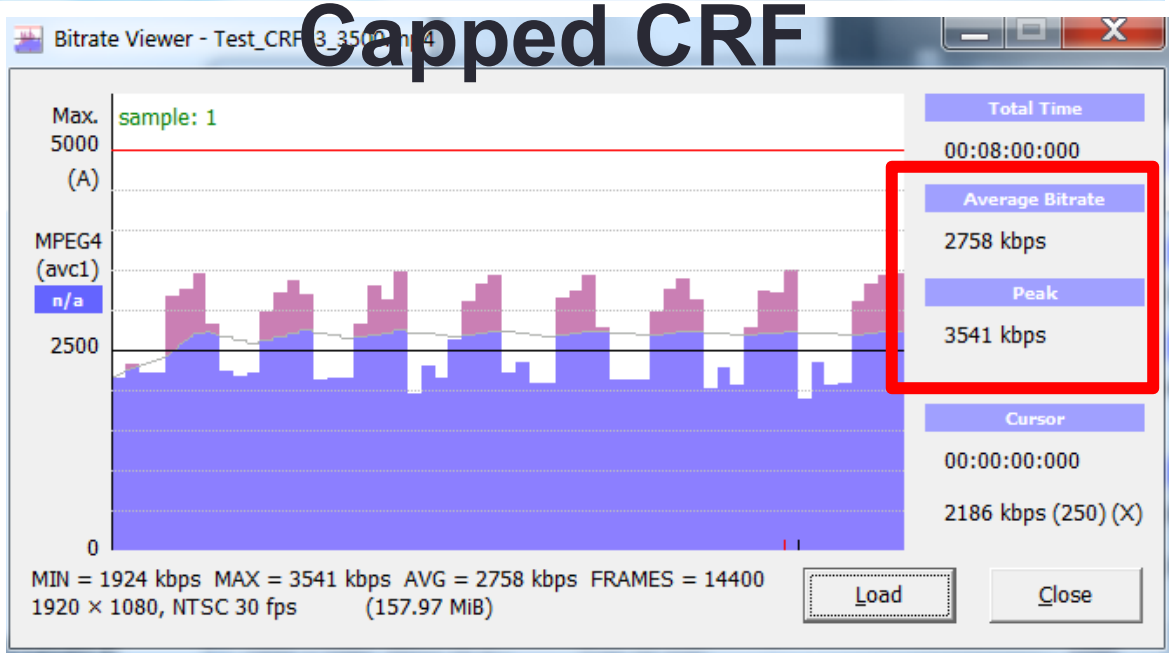
- `-crf` – quality setting
- `-maxrate 3500k` – capping the data rate
- `-bufsize 7000k` – capping the buffer size



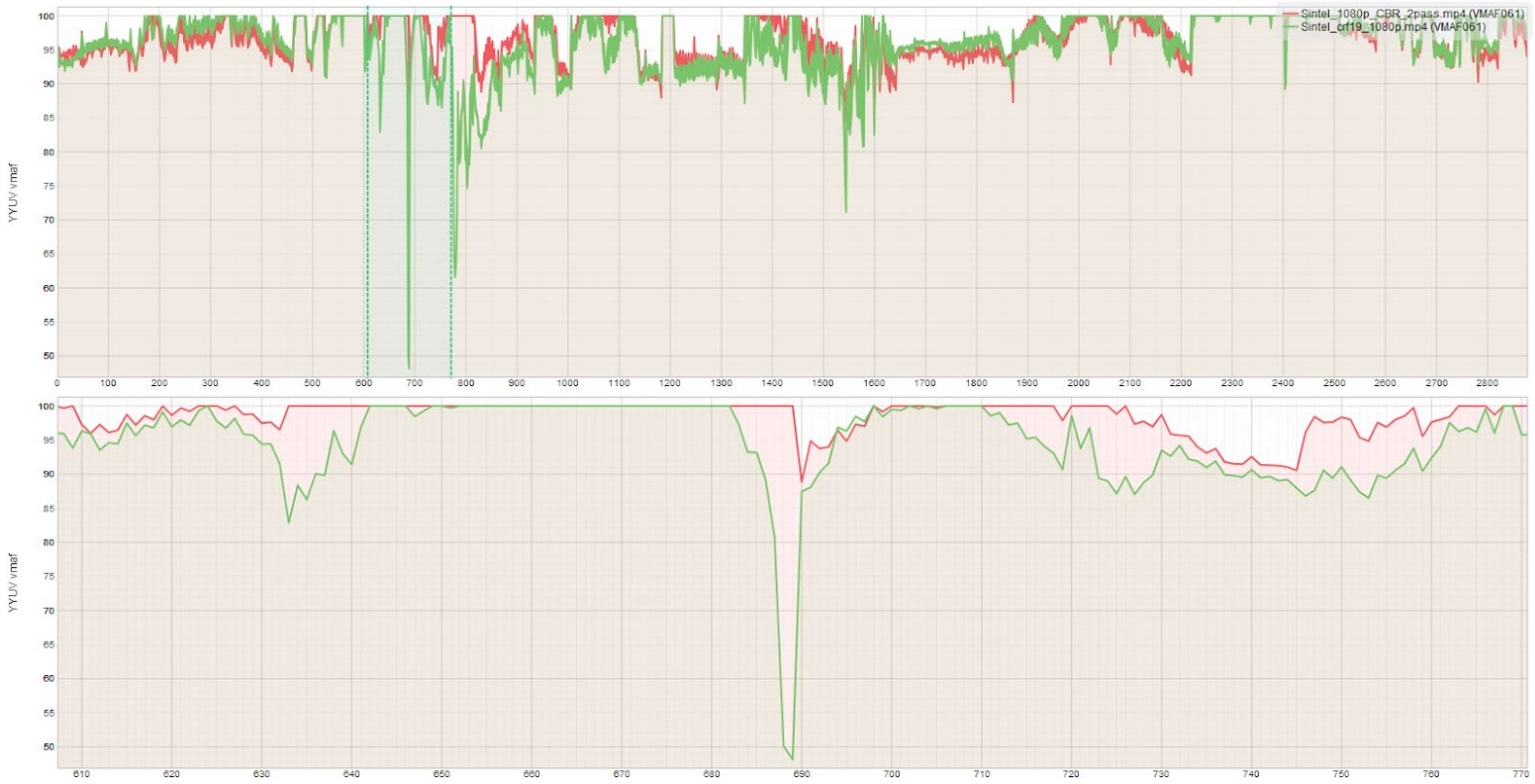
Capped CRF in Action

- CRF on top; capped on the bottom
- Peak data rate
  - CRF – 4455 kbps
  - Capped CRF – 3541 kbps (3500 kbps target)
- Mission accomplished





Caveat: Transient Issues Appear in Many Clips

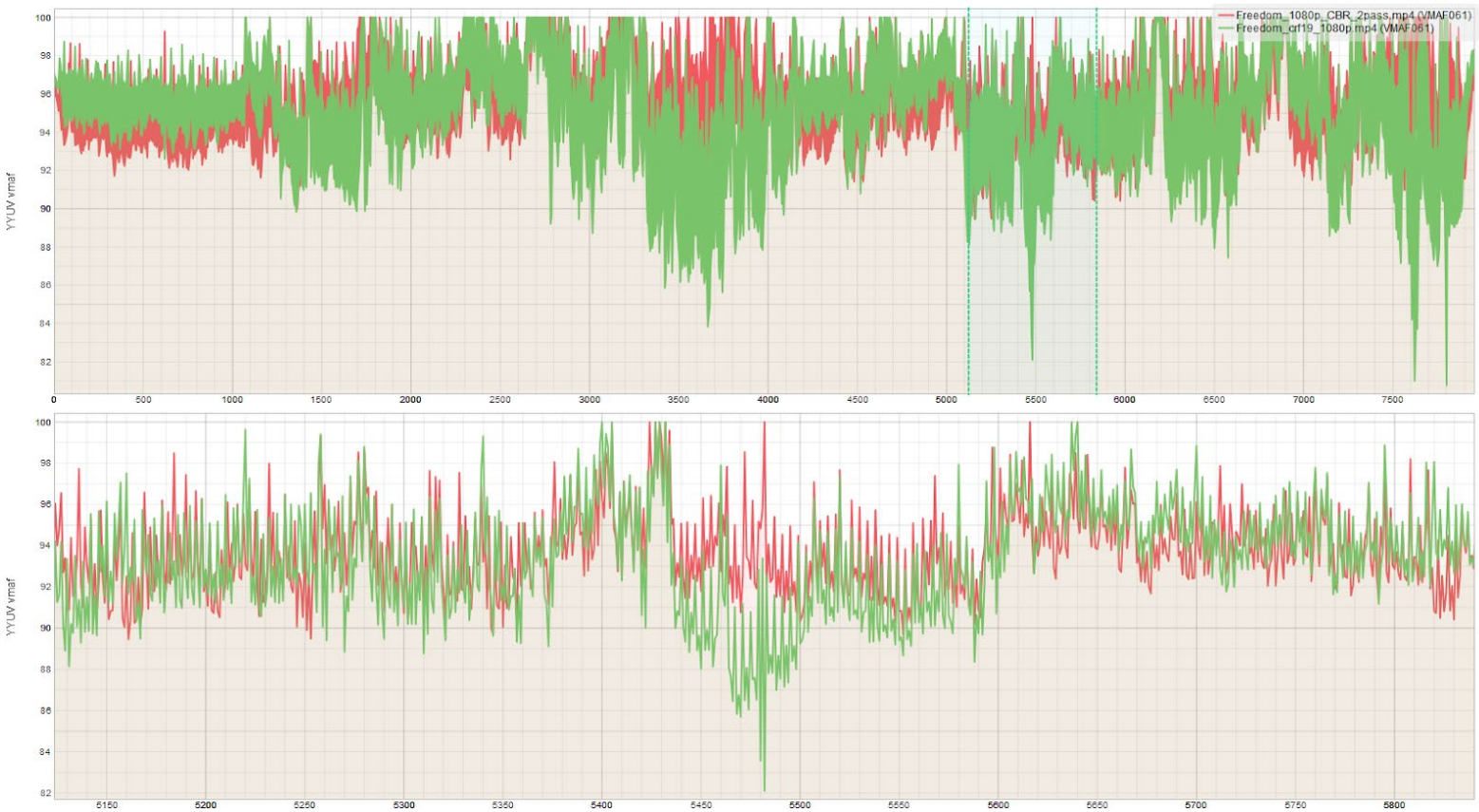


- Sintel clip
  - CRF 19 – green
  - VBR (wrong file name) - red
- Significant score deltas
  - Not really visible


VBR
00:00:28:17


Capped CRF
00:00:28:17

Caveat: Transient Issues Appear in Many Clips

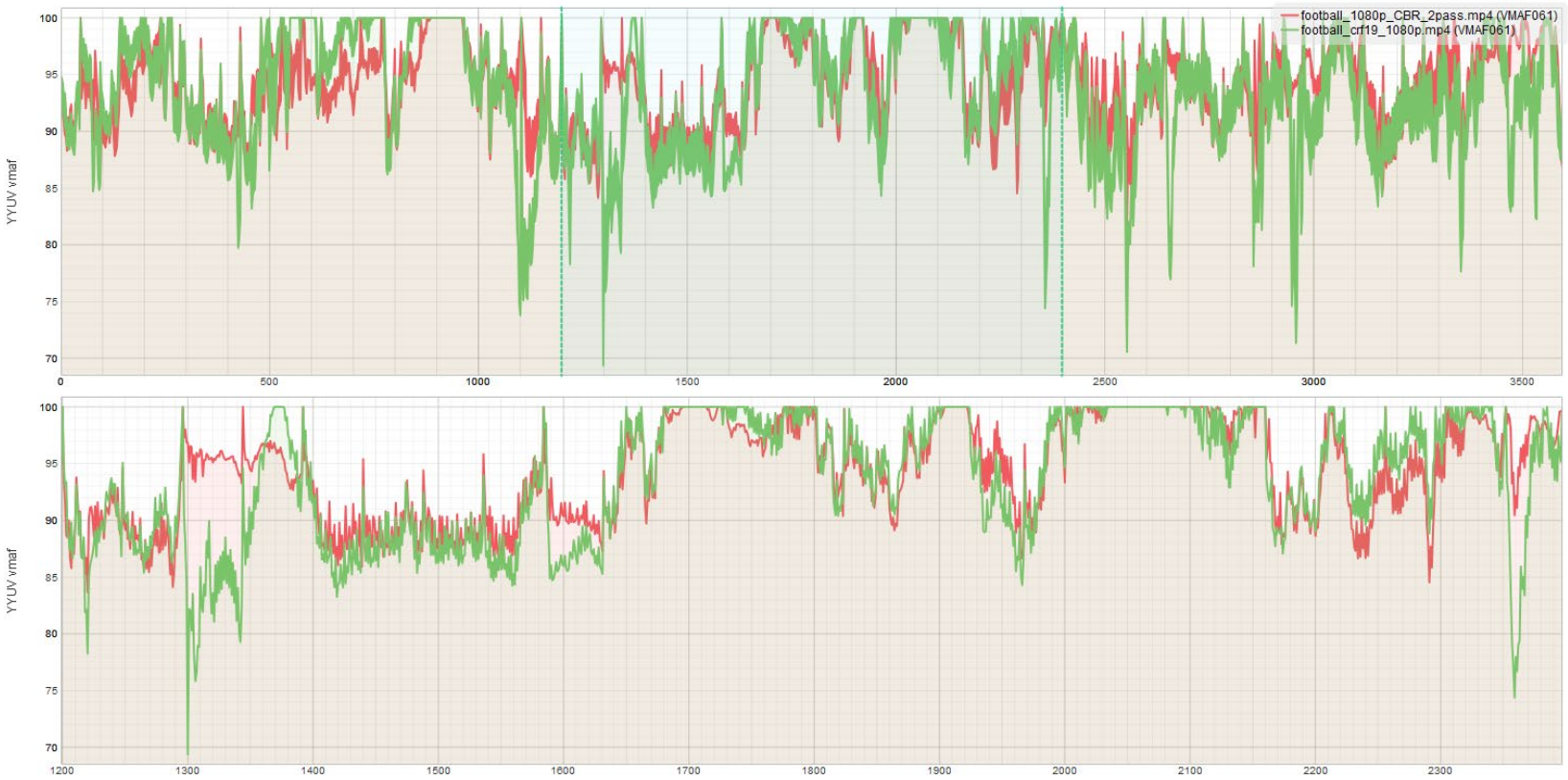


- Freedom clip
  - CRF 19 – green
  - VBR (wrong file name) - red
- Significant score deltas
- Actual quality differences noticeable but not profound


VBR
GRAHAM FEST
01;03;02;28

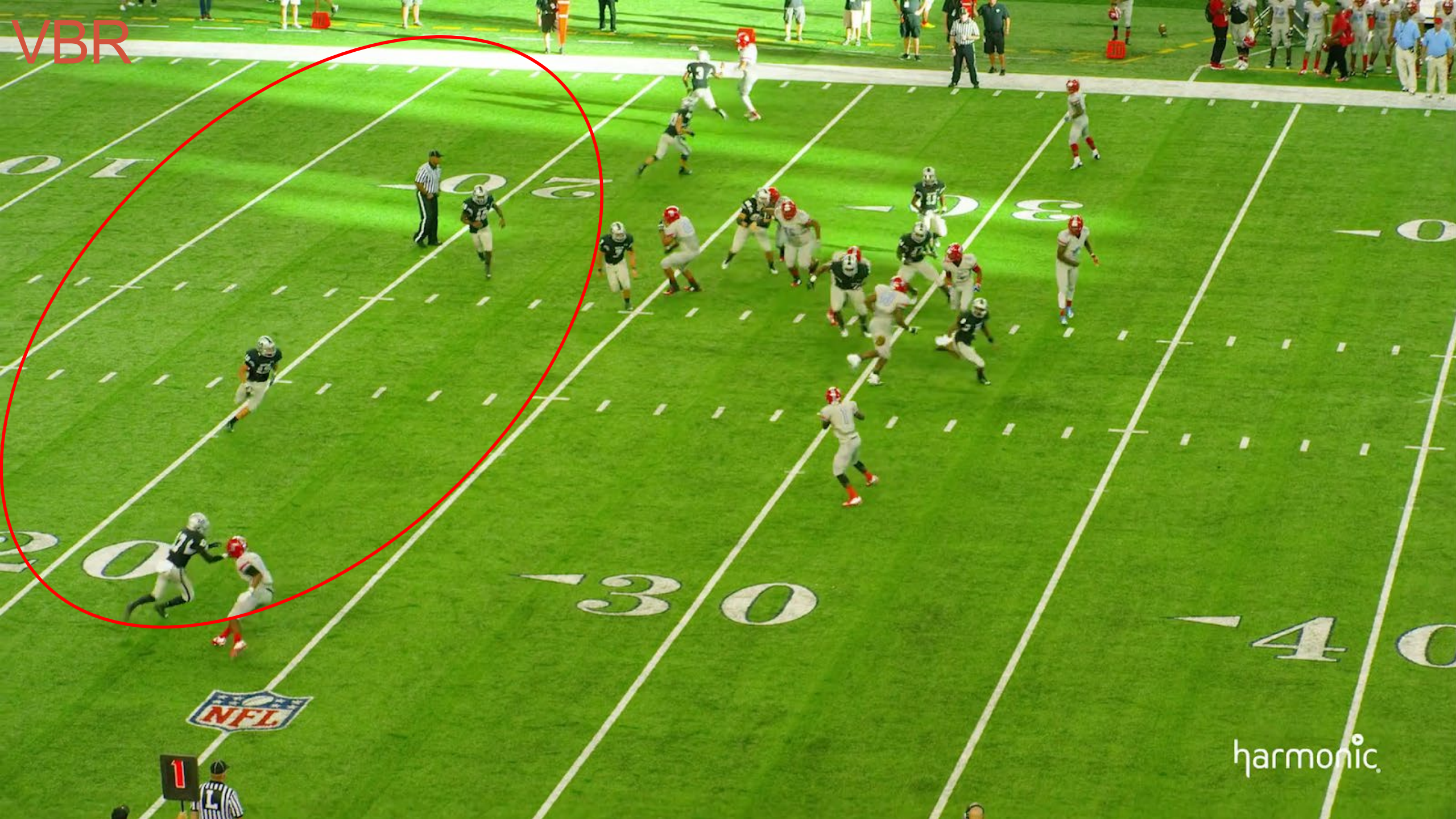

Capped CRF
GRAHAM FEST
01;03;02;28

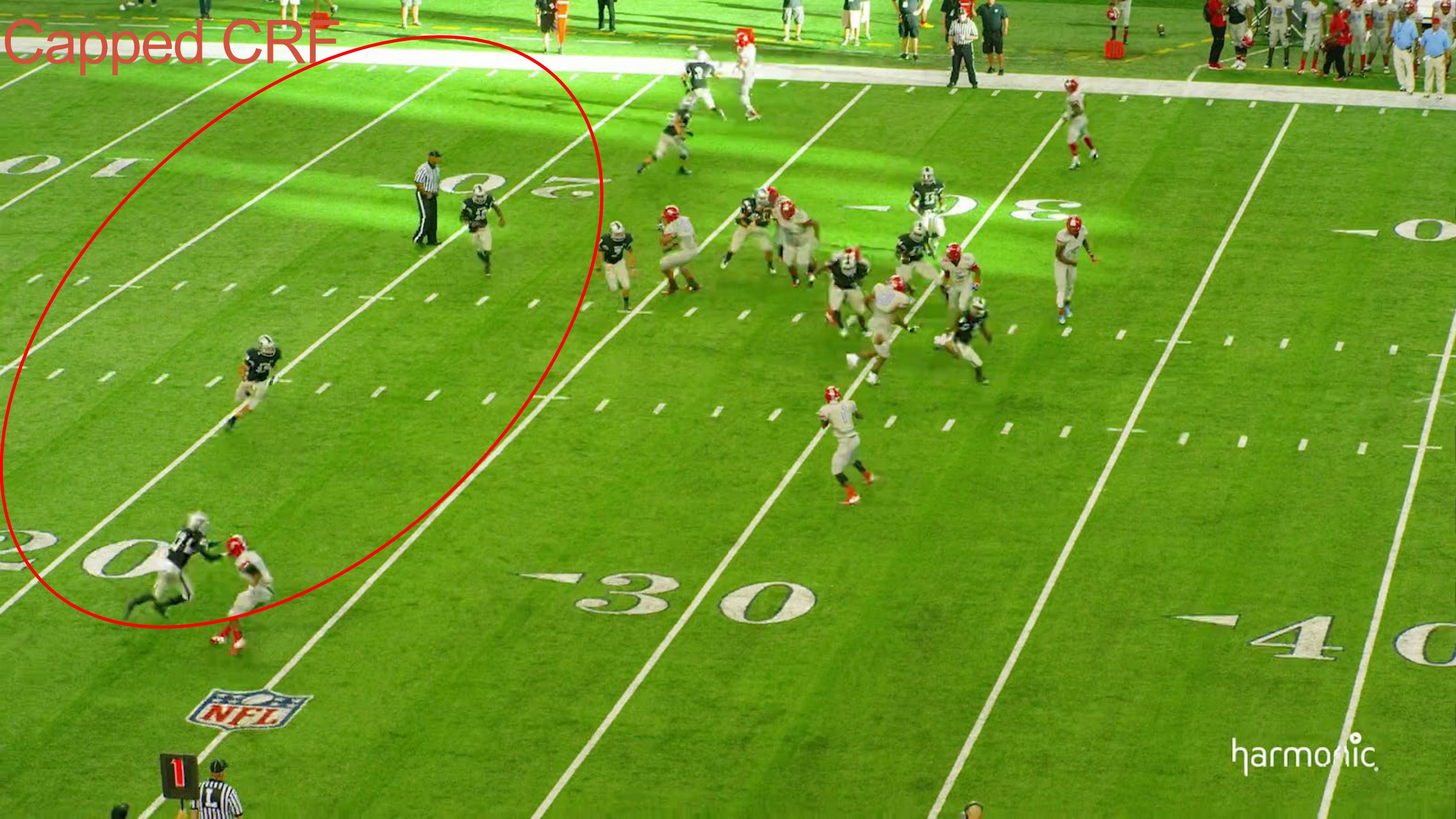
Caveat: Transient Issues Appear in Many Clips



- Football clip
  - CRF 19 – green
  - VBR (wrong file name) - red

- Significant score deltas
- Actual quality differences not that great
- Very transient so probably not noticeable


VBR
NFL
harmonic


Capped CRF
harmonic

Capped CRF Encoding

- Pros:
  - Single pass so efficient
  - Varies quality to match variability for highly cost effective per-title
- Cons
  - Some transient issues like CBR
  - Can have data rate variability

- Capped CRF Users
  - Vimeo
  - JWPlayer (VOD)
  - Brightcove (live)
  - Several large OTT producers
- Bottom line
  - Still a viable DIY technology for those encoding with x264 codec
  - More effective alternatives are available (discuss in per-title section)

Building Your Encoding Ladder with VMAF/CRF

- What's the ceiling
- Compared to YouTube
- How to deploy
  - Capped CRF
  - Multiple encoding ladders

Using Quality Metrics – Finding the Ceiling

- What is the ceiling?
  - The lowest full resolution data rate that delivers acceptable quality
- Finding the ceiling
  - About CRF
  - VMAF correlation
  - Hollywood proof
- Choosing the resolutions

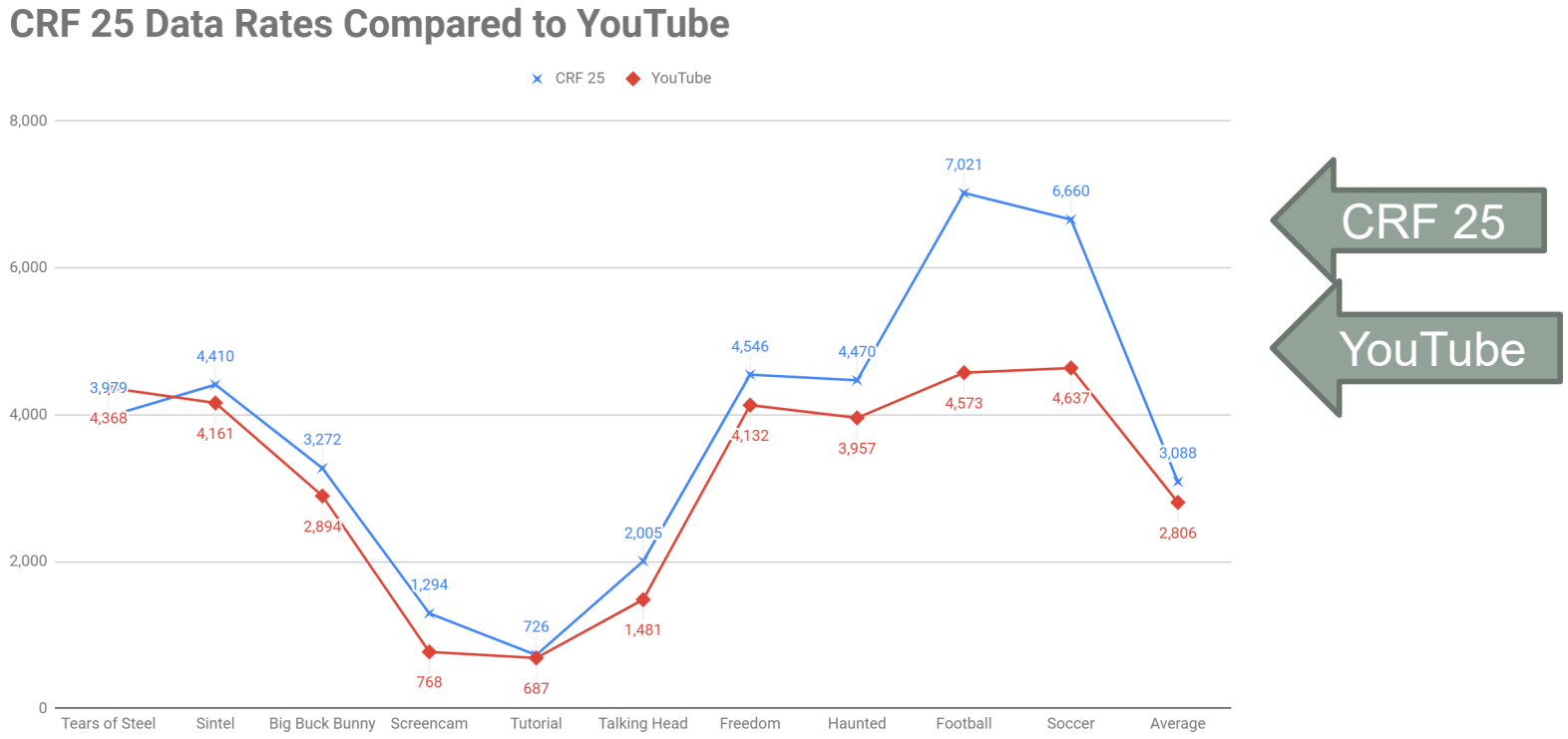
Finding the Optimal Data Rate for 1080p Content (Per-title)

- Compute data rate with CRF 25
  - Encoded 10 files using CRF 25
  - Data rates varied from 725 to 7,021 (~ 1000%)
- Measure VMAF rating
  - Values ranged from 90.40 to 98.46
  - Standard deviation was 2.16 (pretty small)
- Analysis
  - At 2.0 Mbps, a talking head video offers same quality as a football game at 7 Mbps (even lower for synthetic videos)
  - Using the same encoding ladder for both clips is a waste of bandwidth

			CRF	
CRF25 - 1080p	**FPS**	**Description**	**Data Rate**	**VMAF**
Tears of Steel	24	Real world/CG movie	3,979	94.86
Sintel	24	Complex animation	4,410	95.60
Big Buck Bunny	30	Simple animation	3,272	95.94
Screencam	30	Camtasia-based video	1,294	96.00
Tutorial	30	PowerPoint and talking head	726	96.18
Talking Head	30	Simple talking head	2,005	96.18
Freedom	30	Concert footage	4,546	93.78
Haunted	30	DSLR movie-like production	4,470	90.40
Football	30	Football clip	7,021	97.04
Soccer	25	Soccer clip	6,660	98.46
Average			**3,088**	**94.87**
Standard deviation				**2.16**

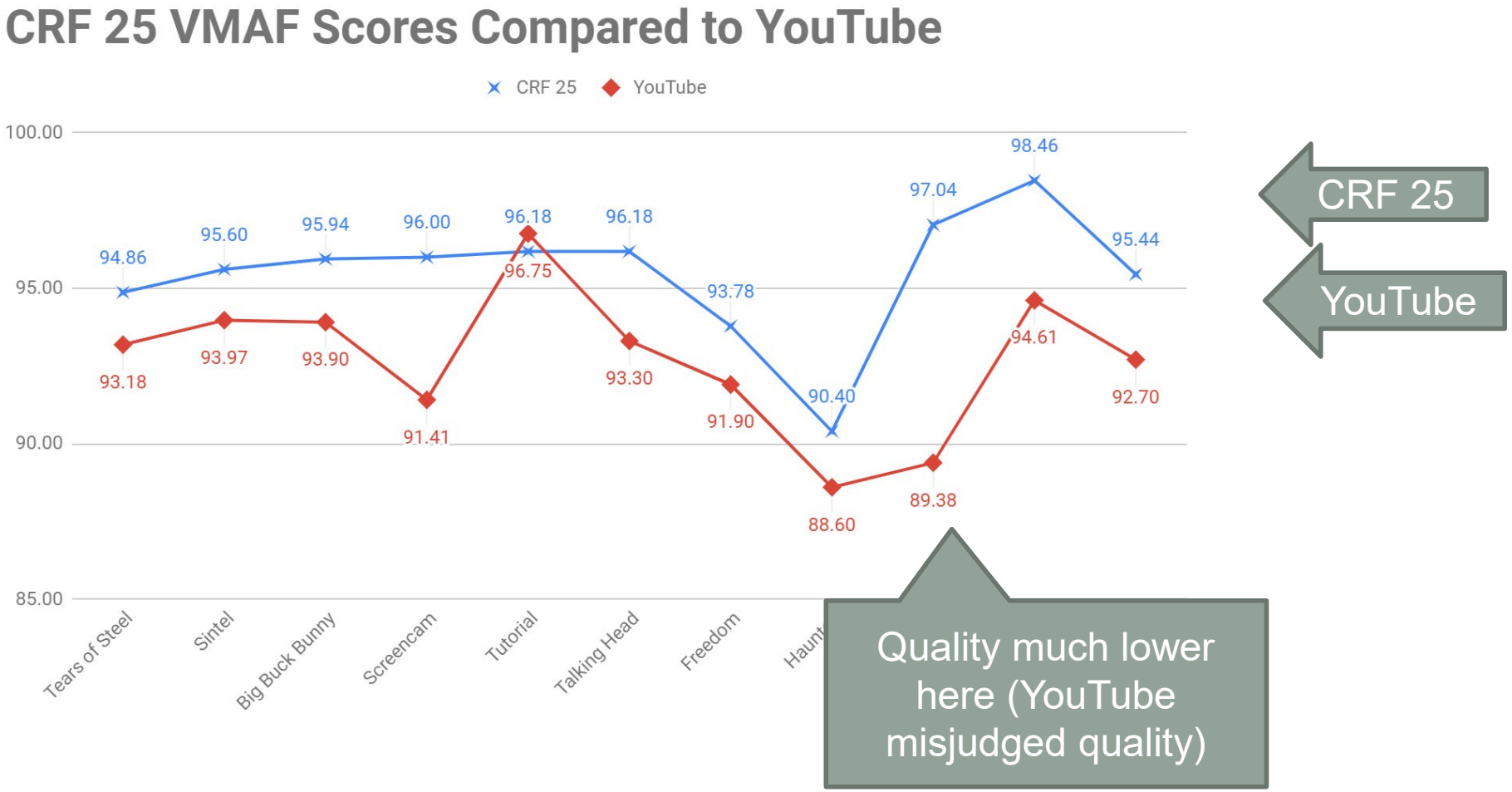
- Conclusion:
  - When computing the appropriate **top data rate** for videos, use VMAF 93
  - CRF 25 with x.264 typically delivers at least VMAF 93

Reality Check: YouTube Comparison - Bitrate



- Upload files to YouTube; measure data rate of 1080p file
  - YouTube uses AI-based per-title optimization
  - Pattern very similar – YouTube views complexity similarly to CRF
- Big deltas for sports Overall, average is very close

Reality Check: YouTube Comparison - VMAF



What's The Point?

			CRF		YouTube		Delta	
CRF25 - 1080p	**FPS**	**Description**	**Data Rate**	**VMAF**	**Data Rate**	**VMAF**	**Data Rate**	**VMAF**
Tears of Steel	24	Real world/CG movie	3,979	94.86	4,368	93.18	389	-1.68
Sintel	24	Complex animation	4,410	95.60	4,161	93.97	-249	-1.64
Big Buck Bunny	30	Simple animation	3,272	95.94	2,894	93.90	-378	-2.03
Screencam	30	Camtasia-based video	1,294	96.00	768	91.41	-526	-4.59
Tutorial	30	PowerPoint and talking head	726	96.18	687	96.75	-39	0.57
Talking Head	30	Simple talking head	2,005	96.18	1,481	93.30	-524	-2.88
Freedom	30	Concert footage	4,546	93.78	4,132	91.90	-414	-1.88
Haunted	30	DSLR movie-like production	4,470	90.40	3,957	88.60	-513	-1.80
Football	30	Football clip	7,021	97.04	4,573	89.38	-2,448	-7.65
Soccer	25	Soccer clip	6,660	98.46	4,637	94.61	-2,023	-3.85
Average			**3,088**	**94.87**	**2,806**	**92.88**	**-282**	**-1.99**
Standard deviation				**2.16**		**2.45**		

- CRF 25 aligns with VMAF 93 to determine acceptable top rung encoding ladder
- CRF 25 also aligns with YouTube's AI-based complexity measure
- CRF 25 is a useful mechanism for gauging complexity and setting the top rung of your encoding ladder

Choosing the Data Rate for Individual Rungs

- Once you know the highest, the rest is just simple math
  - Step 1: Choose highest – VMAF 93
  - Step 2: Choose lowest – slowest speed you want to serve
  - Step 4: fill in the blanks (between 150/200% apart)
    - Apple recommended rule-of-thumb
    - If too far apart, can strand user at lower quality levels
    - If too close, increases encoding cost with minimal benefit

200 kbps

1.67x

360 kbps

1.67x

600 kbps

1.67x

1000 kbps

1. 7x

1700 kbps

1. 6x

2800 kbps

1. 6x

4600 kbps

How Can You Use These Techniques

- Capped CRF
- Category-specific encoding
  - What worked
    - Separate ladder for talk shows and sit coms for major OTT producer
      - Proved that 5 Mbps delivered 93+ VMAF for these types of shows
      - Action shows needed 8 Mbps
    - Online training company
      - Separate ladders for screencam/PowerPoint than real world videos
    - Online bike videos
      - Real world needed 1080p/simple yoga videos fine at 720p

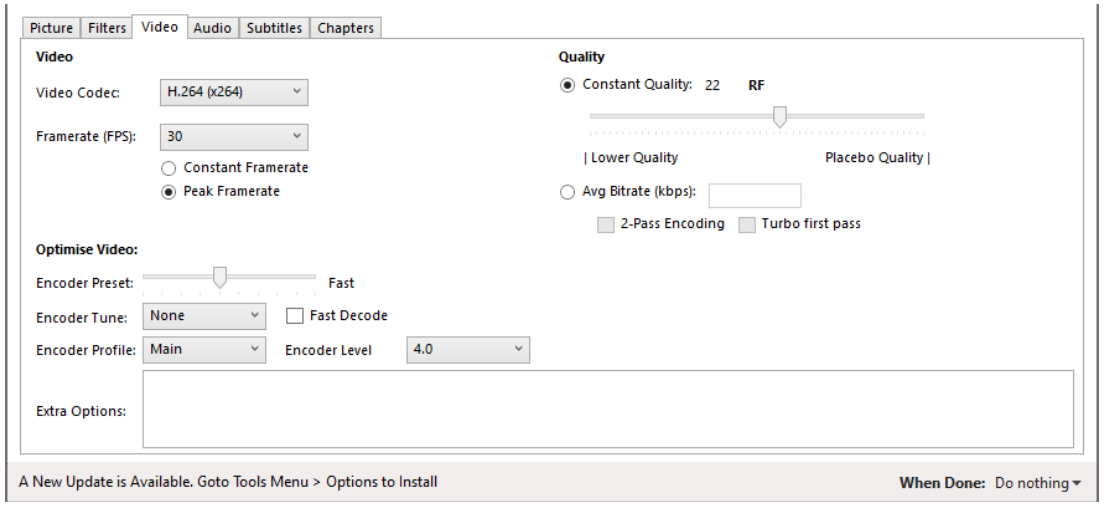
How Can You Use These Techniques

- What didn't work
  - Separate ladders for different kinds of movie (action, etc)
    - Just too much differential within each category
  - Separate ladders for animations vs. movies
    - Just too much differential – Sintel vs. Big Buck Bunny vs SpongeBob

Procedure

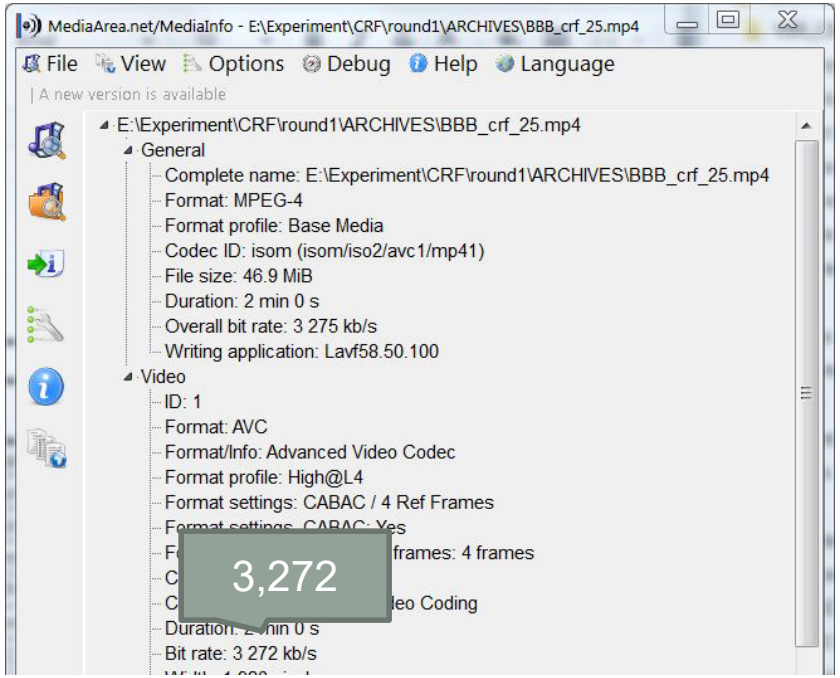
1. CRF 25 encode

```
FFmpeg -i input.mp4  -crf 25 output.mp4
```



Handbrake

2. Check Data Rate in MediaInfo



3. Choose ladder preset

3 Mbps	Data Rate	Max	Buffer
1080p	3,000	6,000	6,000
720p	1,800	3,600	3,600
540p	1,100	2,200	2,200
480p	650	1,300	1,300
360p	400	800	800
270p	250	500	500
180p	150	300	300

1. Manual operation for low volume
2. Scriptable on some platforms

Separate Ladders Triggered by CRF 25 Data Rate

Top	Data Rate	Max	Buffer
1080p	7,000	14,000	14,000
720p	4,200	8,400	8,400
540p	2,500	5,000	5,000
480p	1,500	3,000	3,000
360p	900	1,800	1,800
270p	540	1,080	1,080
180p	325	650	650

6 Mbps	Data Rate	Max	Buffer
1080p	6,000	12,000	12,000
720p	3,600	7,200	7,200
540p	2,200	4,400	4,400
480p	1,300	2,600	2,600
360p	800	1,600	1,600
270p	500	1,000	1,000
180p	300	600	600

5 Mbps	Data Rate	Max	Buffer
1080p	5,000	10,000	10,000
720p	3,000	6,000	6,000
540p	1,800	3,600	3,600
480p	1,000	2,000	2,000
360p	600	1,200	1,200
270p	360	720	720
180p	220	440	440

- Talking head
- Screencam

4 Mbps	Data Rate	Max	Buffer
1080p	4,000	8,000	8,000
720p	2,400	4,800	4,800
540p	1,400	2,800	2,800
480p	840	1,680	
360p	500	1,000	
270p	300	600	
180p	180	360	360

3 Mbps	Data Rate	Max	Buffer
1080p	3,000	6,000	6,000
720p	1,800	3,600	3,600
540p	1,100	2,200	2,200
480p	650	1,300	1,300
360p	400	800	800
270p	250	500	500
180p	150	300	300

2 Mbps	Data Rate	Max	Buffer
1080p	2,000	4,000	4,000
720p	1,200	2,400	2,400
540p	720	1,440	1,440
480p	400	800	800
360p	250	500	500
270p	150	300	300
180p	90	180	180

- PowerPoint video

1 Mbps	Data Rate	Max	Buffer
1080p	1,000	2,000	2,000
720p	600	1,200	1,200
540p	360	720	720
[illegible]p	220	440	440
360p	130	260	260
270p	80	160	160
180p	50	100	100

- 2 Mbps and below may not be needed for many services
- Consider deleting rungs in yellow

Customize For Computer Generated Files

- Simple animated files (cartoons, not heavy CG)
- Synthetic files (PowerPoint/Screencams)

Use higher resolution in rungs below the top

3 Mbps	Data Rate	Max	Buffer
1080p	3,000	6,000	6,000
720p	1,800	3,600	3,600
540p	1,100	2,200	2,200
480p	650	1,300	1,300
360p	400	800	800
270p	250	500	500
180p	150	300	300

3 Mbps	Data Rate	Max	Buffer
1080p	3,000	6,000	6,000
1080p	1,800	3,600	3,600
720p	1,100	2,200	2,200
540p	650	1,300	1,300
540p	400	800	800
360p	250	500	500
270p	150	300	300

Lesson: Introduction to Per-Title Encoding

- What is per-title encoding
- Why is it important
- Where to get it

In the Beginning, Everyone Used TN2224 and it was Good

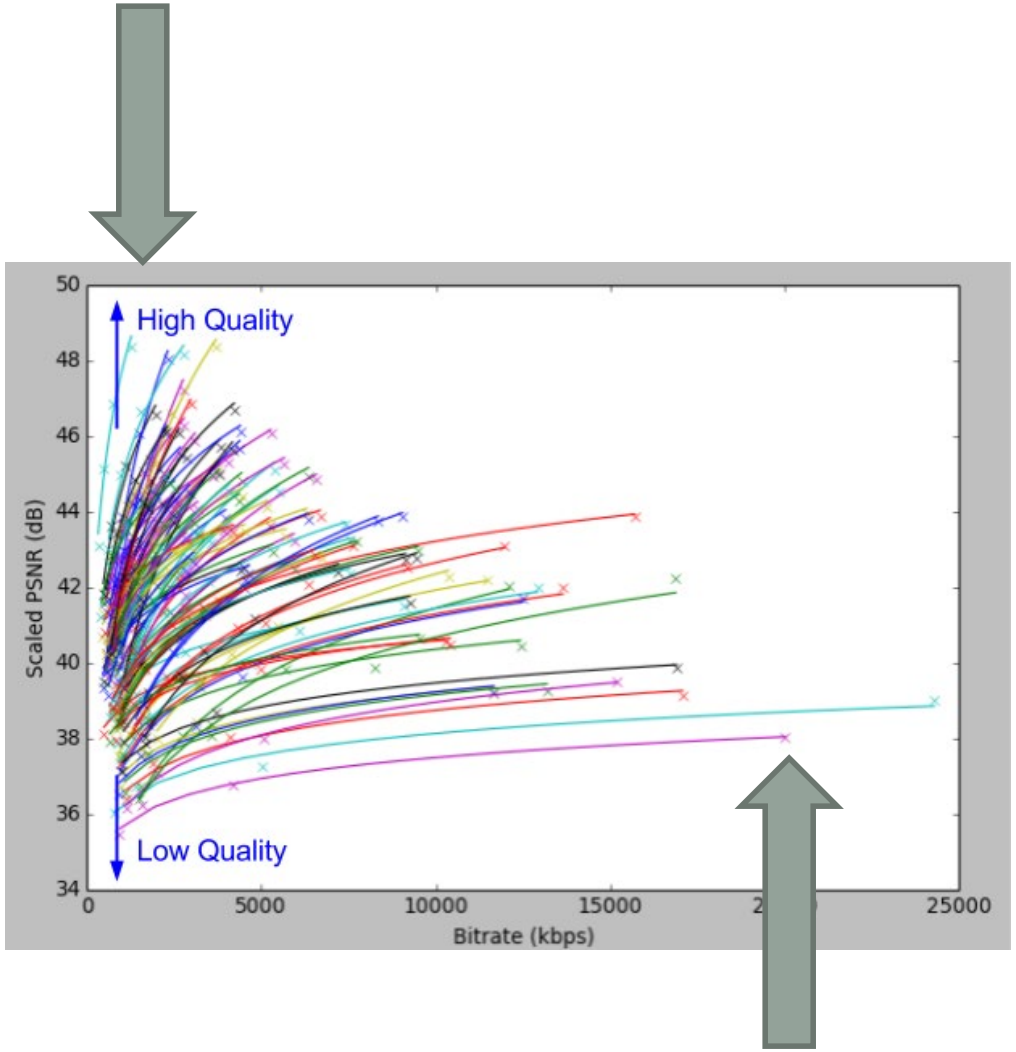
https://blog.beamr.com/2016/03/17/tn2224-is-so-yesterday/

16:9 Aspect Ratio

	Dimensions	Frame Rate *	Total Bit Rate	Audio Bit Rate	Keyframe**
CELL	480x320	na	64	64	na
CELL	416x234	10 to 12	264	64	30 to 36
CELL	480x270	12 to 15	464	64	36 to 45
WIFI	640x360	29.97	664	64	90
WIFI	640x360	29.97	1264	64	90
WIFI	960x540	29.97	1864	64	90
WIFI	960x540	29.97	2564	64	90
WIFI	1280x720	29.97	4564	64	90
WIFI	1280x720	29.97	6564	64	90
WIFI	1920x1080	29.97	8564	64	90

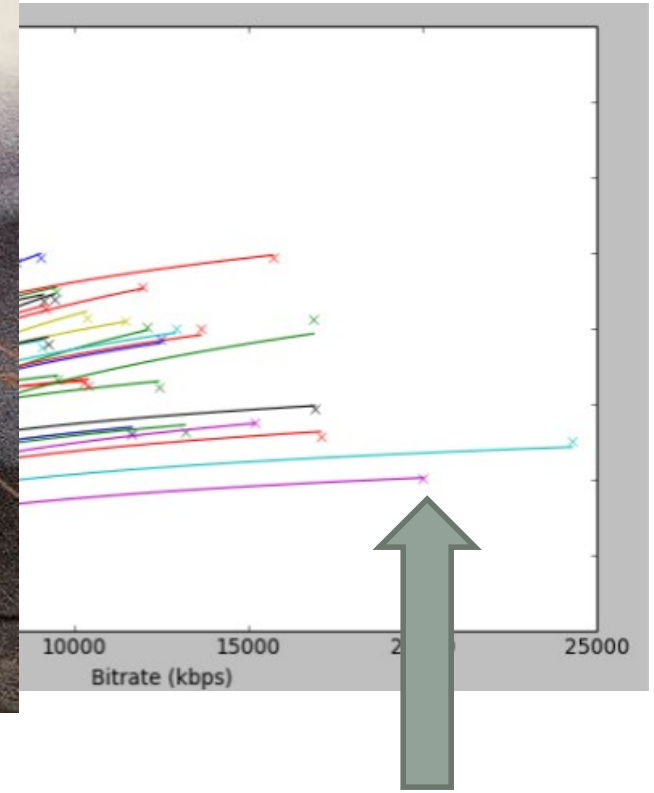
Then, Netflix Invented Per-Title Encoding

- Recognized that:
  - Some files encode to high quality at low bitrates
  - Some files don't achieve high quality even at very high bitrates
  - Makes no sense to use the same encoding ladder for both



Then, Netflix Invented Per-Title Encoding

- Recognized
  - Some files e
    at low bitrate
  - Some files d
    quality even
  - Makes no se
    encoding lac



What is Per-Title Encoding

- Systems that adjust data rate (and resolution and number of rungs in ladder) for complexity and composition
- So, it automates the analysis we just went through, and considers more factors

Why is Per-Title Important?

- If you use a single ladder for all files, you're either wasting bandwidth or producing sub-quality video

Where to Get It (Partial List)

DIY

- Capped CRF
- CRF/manual

Online Video Platforms

- Brightcove

Cloud Vendors

- AWS Elemental
- Bitmovin
- Encoding.com
- Mux
- Zencoder (Brightcove)

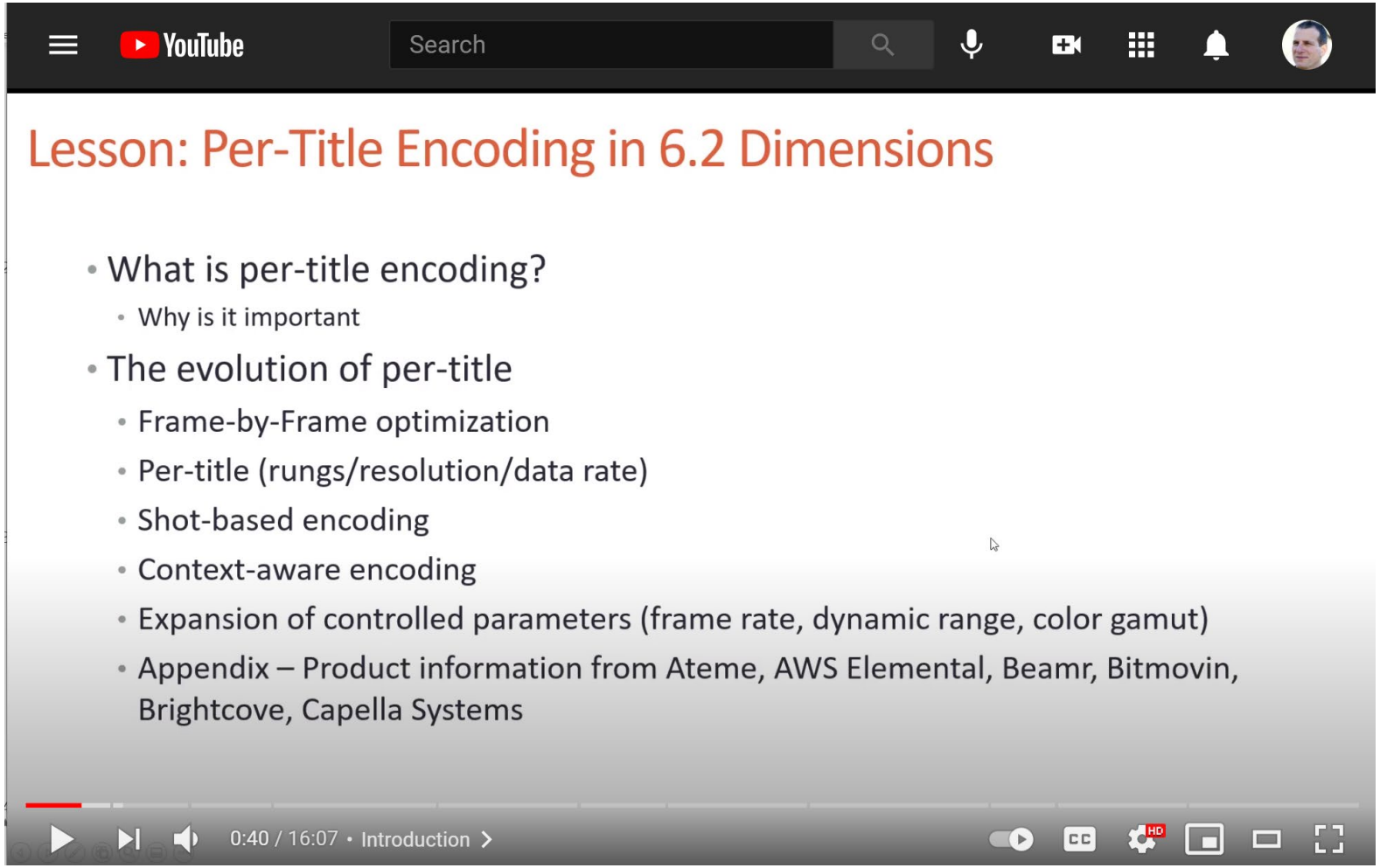
Encoding Vendors

- Ateme
- AWS Elemental
- Capella
- Haivision
- Harmonic

License Technology

- Beamr
- Crunch Media Works

For a Complete Rundown



http://bit.ly/ptenc_jo

For More Information

- Survey of Per-Title Technologies
  - Streaming Media East 2019, handouts and video
  - bit.ly/PT_SME2019_JO
- Choosing the Best Per-Title Encoding Technology
  - White paper written for Bitmovin (winner of SME2019 survey)
  - bit.ly/BM_PT
- Buyers' Guide to Per-Title Encoding - bit.ly/BG_PT
  - From 2019 Streaming Media Sourcebook
-