



Alliance for Open Media

Next Generation, Open-Source Digital Media Technology for
Everyone

AV2 Common Test Condition Latest Test Result Update

Ryan Lei, Meta
2025 Fall VQEG Meeting, Dec 15-19, 2025

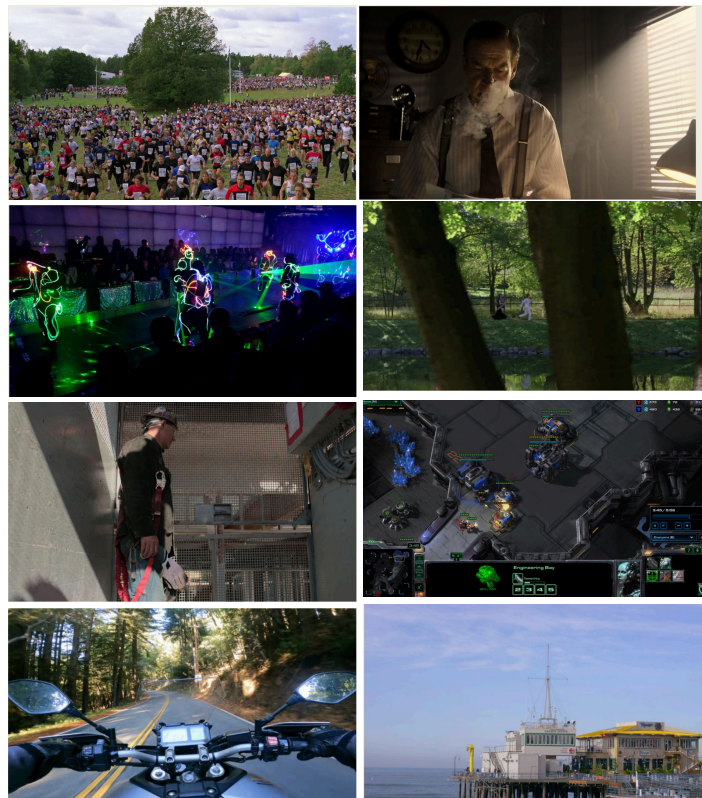
AOM Testing Subgroup Activities

- Define Common Test Conditions for evaluating coding gain of AV2 tools
 - Test sequences, encoding configurations, etc.
 - CTC Test Sequences can be found at: https://media.xiph.org/video/aomctc/test_set/
 - Quality and performance metrics.
 - Test script and infrastructure.
 - Latest AOM CTC v8.0 ([CWG-F384](#)).
 - CTC extended to cover 422/444 extended color format ([CWG-F169](#)):
- Conduct regular test and tools on/off test after each anchor release.
 - [AVM v11.0 anchor](#) was released in September 2025.
 - [AVM v12.0 anchor](#) was released in November 2025.
- Define and conduct subjective test.

AOM CTC Test Sequences

Class	Format
A1	3840x2160, 10bit, 4:2:0
A2	1920x1080, 8/10 bit, 4:2:0
A3	1280x720, 8/10 bit, 4:2:0
A4	640x360, 8bit, 4:2:0
A5	480x270, 8/10 bit, 4:2:0
B1	1920x1080, 8/10 bit, 4:2:0, Synthetic content for gaming and animation
B2	1920x1080, 8/10 bit, 4:2:0, Synthetic content for screen sharing
F1	Still Image high resolution upto 8K, 8bit, 4:2:0
F2	Still Image medium resolution upto 2K, 8bit, 4:2:0
G1	3840x2160, 10bit, 4:2:0, HDR videos in BT.2100 color space with PQ transfer function.
G2	1920x1080, 10bit, 4:2:0, HDR videos in BT.2100 color space with PQ transfer function.
E	User Generated Content (UGC), 2K, 4K, 8bit, 4:2:0

91 videos, 51 images



AOM CTC Encoding Configurations

- **Focusing on normative coding tools**
 - Disabled Key Frame Filtering, adaptive QP modulation, Two Pass encoding, etc.
 - Different tiling configurations specified to speedup encoding test.
- **All Intra (AI)**
 - Mainly used for encoding Key frames and still images.
 - Encoding first 30 frames from all classes as Intra frames
- **Random Access (RA)**
 - 130 frames, closed-GOP, GOP Size = 65, sub GOP = 16, 5 Hierarchical Layers
 - Single pass encoding with constant quality mode.
 - Fixed QP offset applied for frames at each layer
- **Low Delay (LD)**
 - Encoding 130 frames (with only 1 Key frame) without future reference frames
- **Adaptive Streaming (AS)**
 - 4K sequences in A1 class are downscaled to 5 different resolutions.
 - Similar encoding configuration as Random Access is used for encoding.

Configuration	Qindex values
Still Image	60, 85, 110, 135, 160, 185
All Intra (AI)	85, 110, 135, 160, 185, 210
Random Access (RA)	110, 135, 160, 185, 210, 235
Low Delay (LD)	110, 135, 160, 185, 210, 235
Adaptive Stream (AS)	110, 135, 160, 185, 210, 235

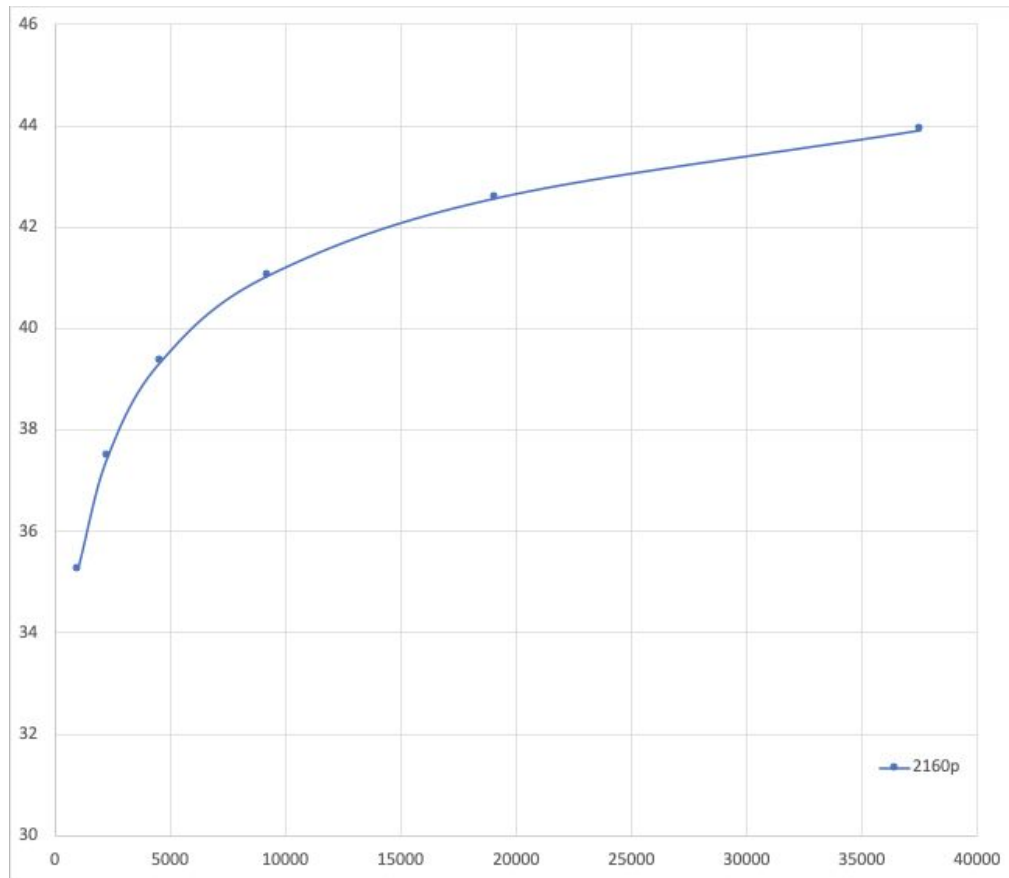
*qindex range is [0, 255]

Hierarchical Layer	Qindex values
1	61 (Key Frame), 79 (Inter)
2	99
3	105
4	108
5	110

Example of QP modulation in RA

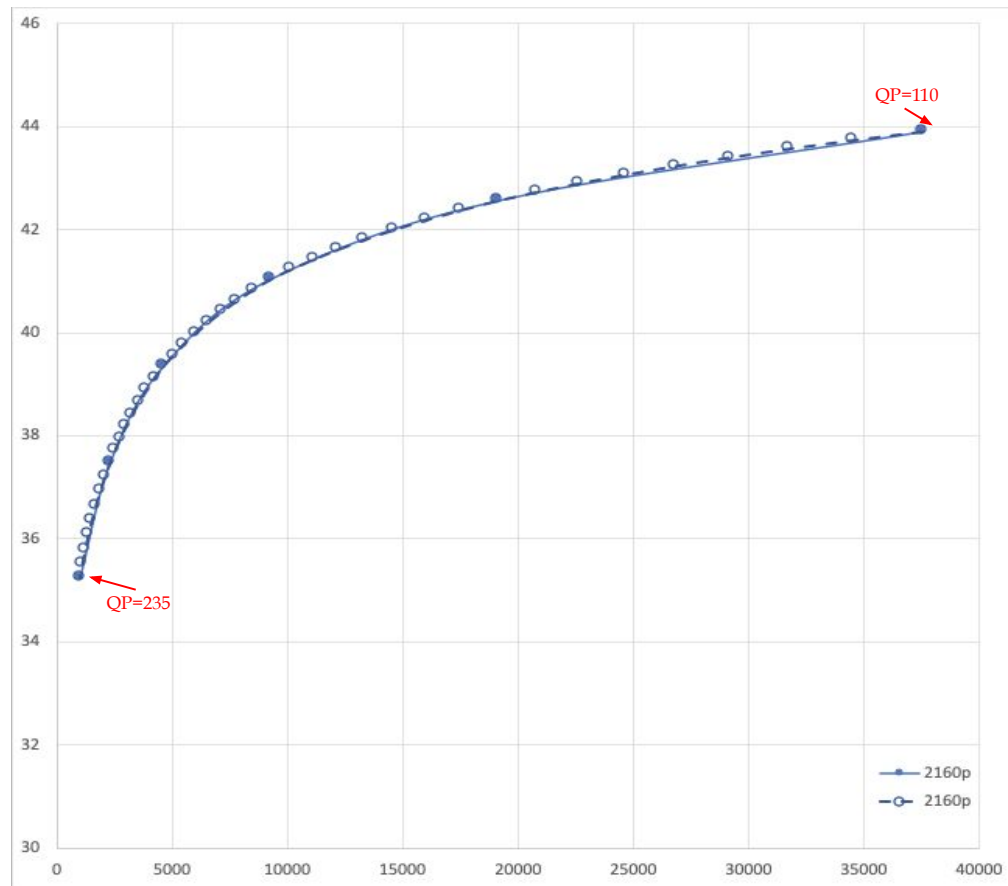
Adaptive Streaming Configuration

- 6 Resolutions are used for each 4K sequence in the A1 class.
- Same 6 QPs are used for encoding following Random Access configuration.
- Downscaled (with Lanczos 5) videos are encoded, decoded and upscaled back to 4K resolution
- Quality metrics are calculated against 4K resolution.



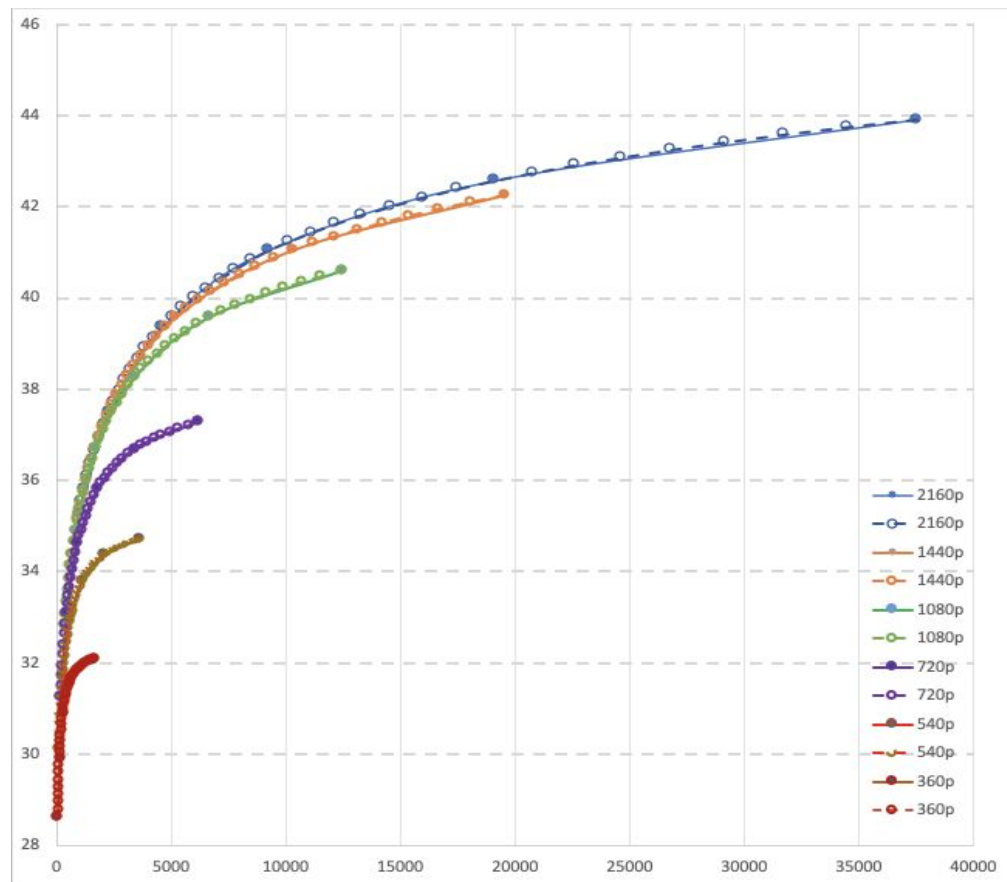
Adaptive Streaming Configuration

- Bilinear Interpolated RD points are generated based on bit rate in log domain
- Totally 41 points are generated for each resolution.



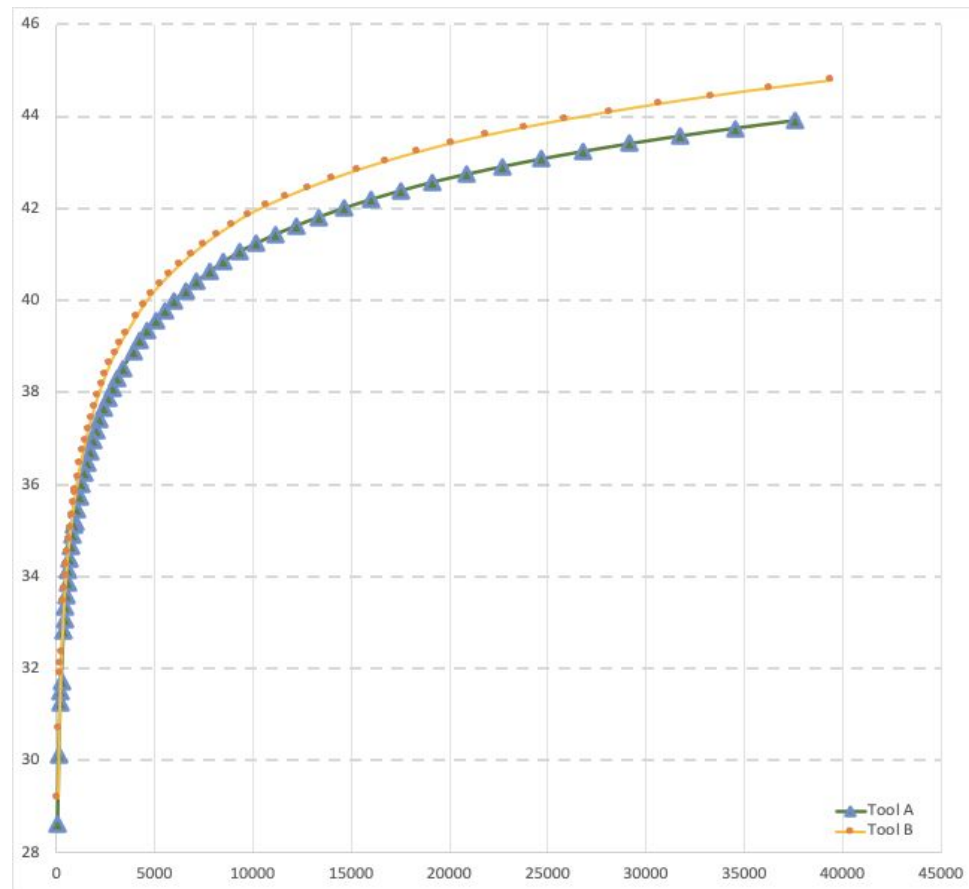
Adaptive Streaming Configuration

- RD points for all resolutions and QPs are used to calculate the convex hull



Adaptive Streaming Configuration

- BDRATE between two convex hulls are calculated to evaluate quality difference between coding tools.

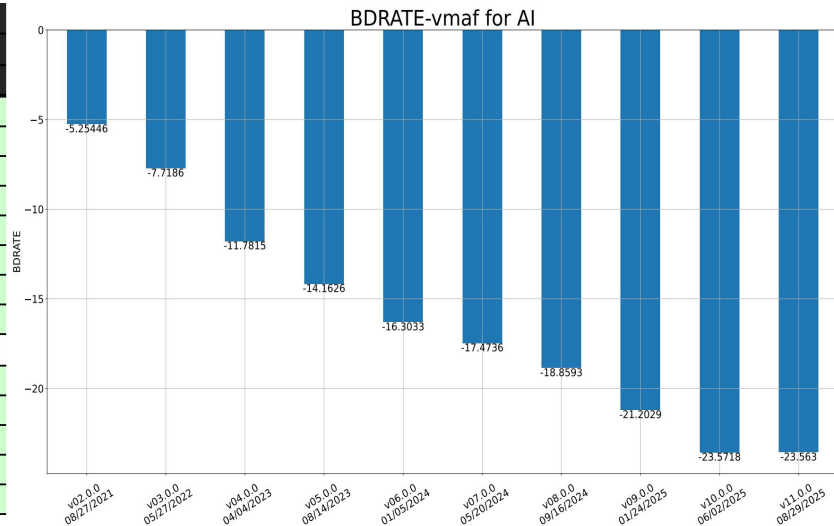


Quality Metrics

- PSNR-Y/U/V, SSIM-Y, MS-SSIM-Y, VMAF, VMAF-NEG, PSNR HVS, DE2000, APSNR-Y/U/V are calculated using [VMAF tool](#) contributed by Netflix. “--aom_ctc v6.0” is used to specify the latest implementation of quality metrics.
 - [VMAF-NEG](#): for codec evaluation purposes to assess the pure effect coming from compression.
 - PSNR-YUV is calculated from PSNR Y, U, V with 14:1:1 weighting factor.
 - APSNR-YUV is calculated from MSE Y, U, V with 4:1:1 weighting factor.
- Special handling for VMAF in BDRATE calculation
 - To solve the problem with VMAF non-monotonicity in a flat (saturated) region of the curve, if VMAF non-monotonicity happens at VMAF value 99.5 or above, the non-monotonic value and the values corresponding to bitrates higher than the non-monotonic value are excluded from the BD-rate calculation.
- [CAMBI \(Contrast Aware Multiscale Banding Index\)](#) metric for measuring banding artifacts is included.
- BDRATE Calculation
 - 6 QPs values are used for encoding. BDRATE for low/medium/high QPs ranges are also calculated.
 - PCHIP interpolation is used to make sure interpolated curve cross the actual measurement.

AVM v11.0 Test Result: All Intra

	Averages Over					
	PSNR		SSIM	MS-SSIM	VMAF	nVMAF
	Y	YUV	Y	Y	Y	Y
Summary						
Class A1_4K	-18.38%	-20.12%	-15.52%	-17.35%	-22.29%	-21.43%
Class A2_2K	-16.87%	-18.67%	-15.28%	-15.74%	-20.11%	-19.35%
Class A3_720p	-18.89%	-20.59%	-17.52%	-17.70%	-22.47%	-21.44%
Class A4_360p	-15.31%	-18.33%	-16.21%	-14.85%	-18.10%	-17.12%
Class A5_270p	-13.84%	-15.20%	-14.03%	-14.23%	-20.52%	-19.46%
Class B1_SYN	-21.27%	-23.13%	-19.86%	-20.44%	-23.87%	-22.84%
Overall Mandatory w/o B2	-17.79%	-19.68%	-16.48%	-16.91%	-21.27%	-20.36%
Class B2_SCC	-40.63%	-42.42%	-40.84%	-42.03%	-40.39%	-39.03%
Class HDR1_4K	-19.21%	-23.51%	-15.60%	-17.43%	-22.52%	-22.20%
Class HDR2_2K	-20.62%	-24.31%	-18.87%	-19.39%	-22.10%	-21.62%
Overall HDR	-20.10%	-24.02%	-17.68%	-18.67%	-22.25%	-21.83%
Class E	-16.82%	-18.65%	-15.88%	-15.86%	-22.95%	-22.22%
Overall	-20.16%	-22.27%	-18.93%	-19.42%	-23.57%	-22.71%



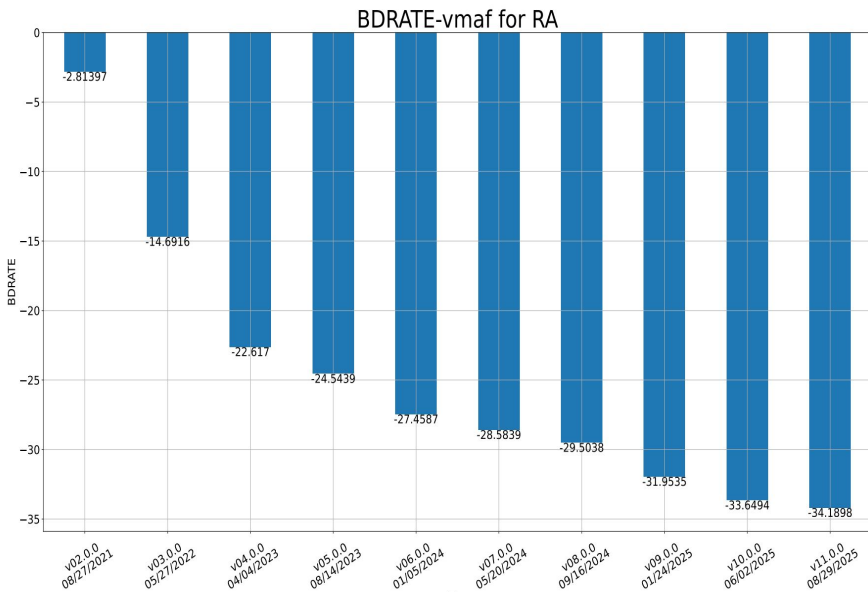
¹ Anchor is [research-alt-v1-anchor_r3.0](#) release tag, test is [research-v11.0.0-rc1](#) release tag.

² PSNR-YUV = (14 * PSNR-Y + PSNR-U + PSNR-V) / 16

³ Encoded with single tile

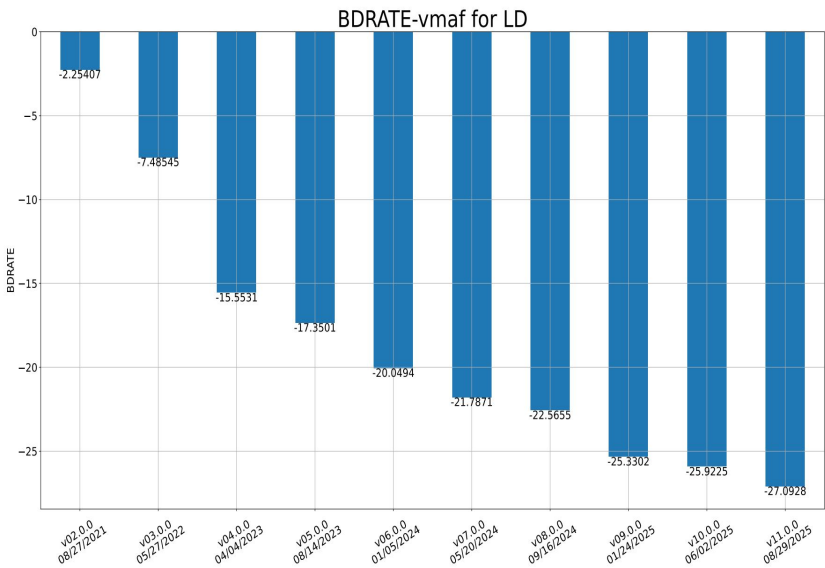
AVM v11.0 Test Result: Random Access

	Averages Over					
	PSNR		SSIM	MS-SSIM	VMAF	nVMAF
	Y	YUV	Y	Y	Y	Y
Summary						
Class A1_4K	-29.30%	-30.73%	-27.32%	-28.19%	-35.52%	-34.31%
Class A2_2K	-27.15%	-28.40%	-25.28%	-25.84%	-33.15%	-32.03%
Class A3_720p	-28.68%	-30.12%	-27.21%	-27.52%	-32.62%	-31.28%
Class A4_360p	-27.05%	-29.00%	-25.25%	-25.42%	-31.99%	-31.75%
Class A5_270p	-28.31%	-28.87%	-26.87%	-28.01%	-34.78%	-34.20%
Class B1_SYN	-26.24%	-27.96%	-26.56%	-26.78%	-32.02%	-30.87%
Overall Mandatory w/o B2	-27.59%	-29.01%	-26.20%	-26.71%	-33.21%	-32.17%
Class B2_SCC	-39.03%	-40.33%	-40.45%	-40.79%	-42.93%	-41.62%
Class HDR1_4K	-32.72%	-35.15%	-30.87%	-30.68%	-43.19%	-41.60%
Class HDR2_2K	-28.26%	-30.96%	-27.08%	-27.42%	-34.60%	-33.66%
Overall HDR	-29.88%	-32.48%	-28.46%	-28.61%	-37.73%	-36.55%
Class E	-24.37%	-26.11%	-23.19%	-23.13%	-30.31%	-29.30%
Overall	-28.43%	-30.04%	-27.36%	-27.70%	-34.20%	-33.13%



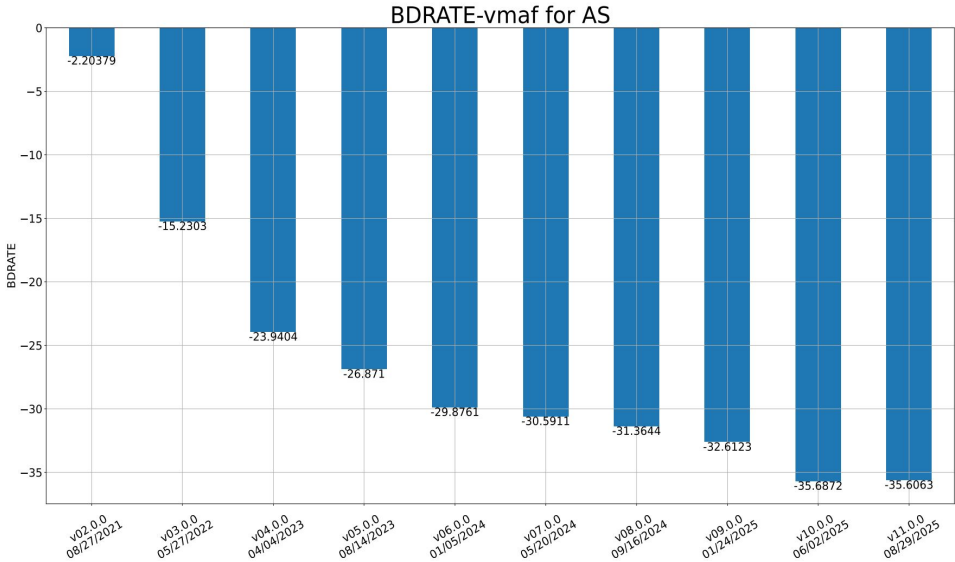
AVM v11.0 Test Result: Low Delay

	Summary	Averages Over					
		PSNR		SSIM	MS-SSIM	VMAF	nVMAF
LD		Y	YUV	Y	Y	Y	Y
	Class A1_4K						
	Class A2_2K	-20.88%	-22.32%	-19.16%	-19.31%	-23.97%	-23.50%
	Class A3_720p	-21.48%	-23.04%	-19.68%	-19.83%	-22.65%	-22.11%
	Class A4_360p	-23.04%	-25.10%	-22.09%	-21.32%	-23.86%	-24.61%
	Class A5_270p	-20.42%	-21.29%	-18.79%	-19.50%	-25.49%	-24.72%
	Class B1_SYN	-20.73%	-22.69%	-20.31%	-20.44%	-23.31%	-22.83%
	Overall Mandatory w/o B2	-21.19%	-22.79%	-19.83%	-19.91%	-23.72%	-23.37%
	Class B2_SCC	-40.16%	-41.29%	-41.45%	-41.47%	-44.79%	-43.13%



AVM v11.0 Test Result: Adaptive Streaming

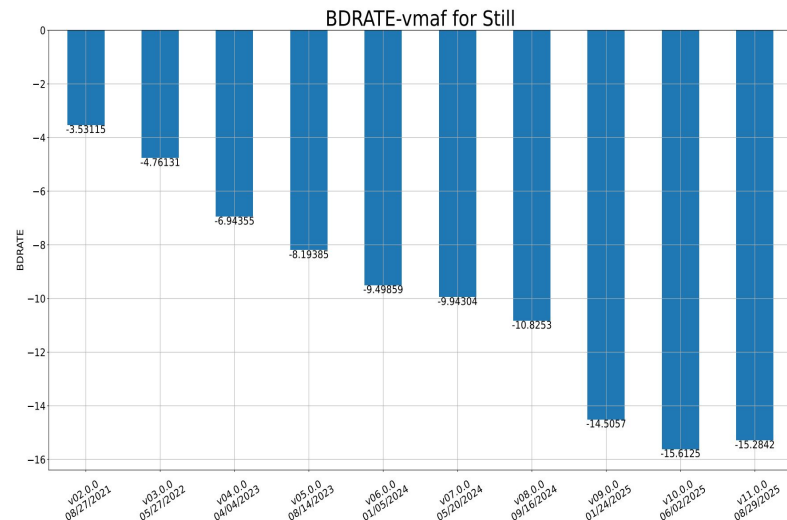
Sequence				
				PSNR
		Y	Y	YUV
Average	High	-24.33%	-30.43%	-31.46%
	Med	-29.03%		
	Low	-39.19%		



Sequence							
	SSIM CH		MS-SSIM CH		VMAF CH		
	SSIM		MS-SSIM		VMAF		
Average	High	Range	Y	Range	Y	Range	Y
	Med	-21.98%	-25.90%	-22.72%	-27.97%	-22.70%	-36.05%
	Low	-25.00%		-27.43%		-36.08%	
		-36.07%		-38.01%		-44.12%	

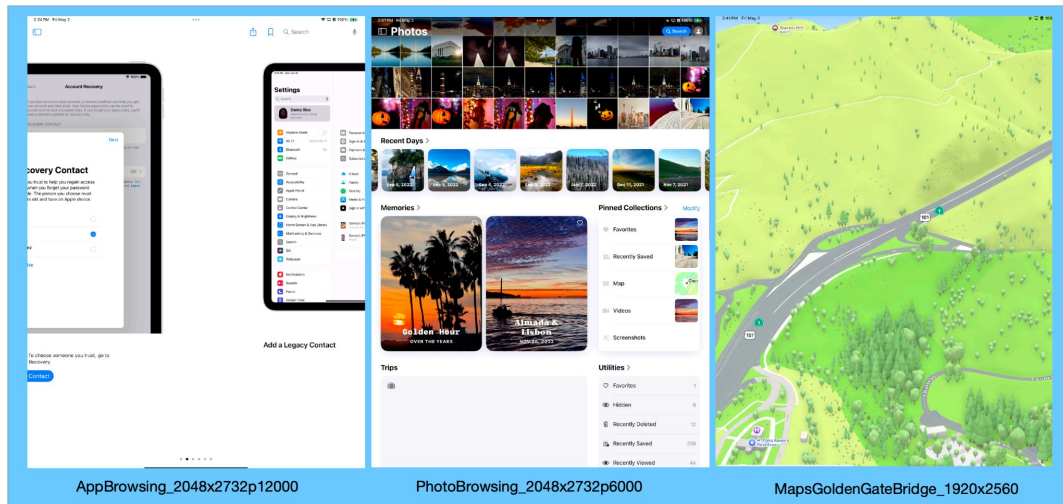
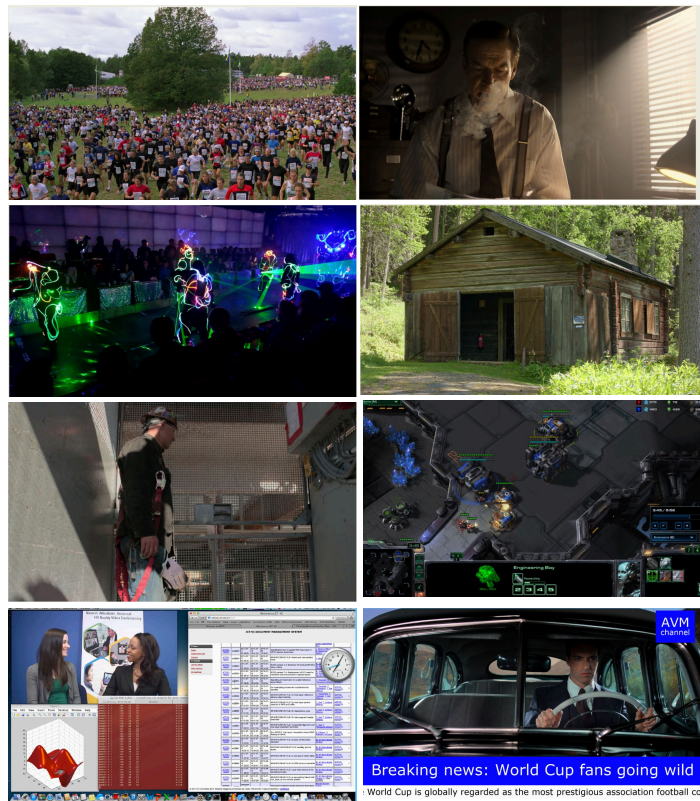
AVM v11.0 Test Result: Still Images

Summary	Averages Over					
	PSNR		SSIM	MS-SSIM	VMAF	nVMAF
	Y	YUV	Y	Y	Y	Y
F1 High Resolutions	-11.56%	-13.47%	-11.12%	-11.98%	-15.27%	-14.51%
F2 Medium Resolution	-12.79%	-14.45%	-13.15%	-13.39%	-15.30%	-14.87%



AV2 Extended Color Format (ECF) Test Conditions

- ECF CTC ([CWG-F169](#))
- **Test Sequences:** 6 classes (ECF1, ECF2, ECF3, ECF4, ECF5, ECF6)
 - 34 sequences: 21 YUV4:4:4 sequences and 13 YUV 4:2:2 sequences
 - One SCC class (ECF6)
 - Resolution: 640x360 to 4k
 - Bit-depth: 8 and 10
 - SDR/HDR/RGB/YCoCg
 - Frame rate: Min 15, Max 120. Mainly 30, 50, 60



AVM Test Result for 422/444 Extended Color Format

- Anchor: commit 0db972de0ce
- Test: AVM main branch commit: 1f8a1e82 (post AVM v10.0 release)
- PSNR-YUV is calculated from PSNR Y, U, V with 6:1:1 weighting factor

	Summary	PSNR		SSIM	MS-SSIM	VMAF
		Y	YUV	Y	Y	Y
Intra	Class ECF1_444_SDR	-19.38%	-19.62%	-19.59%	-18.71%	-22.06%
	Class ECF2_422_SDR	-15.82%	-17.29%	-15.07%	-14.98%	-18.67%
	Class ECF3_444_HDR	-19.48%	-22.80%	-16.57%	-17.69%	-21.46%
	Class ECF4_422_HDR	-22.59%	-24.60%	-20.05%	-21.64%	-24.94%
	Class ECF5_444_YCgCo	-15.87%	-16.15%	-14.70%	-15.26%	-20.92%
	Class ECF6_444_SCC	-31.45%	-32.55%	-28.87%	-31.67%	-34.47%
	Overall Mandatory 444	-18.91%	-20.03%	-18.03%	-17.93%	-21.72%
	Overall Mandatory 422	-18.08%	-19.73%	-16.73%	-17.20%	-20.76%
	Overall	-21.27%	-22.57%	-19.86%	-20.58%	-24.08%

RA	Class ECF1_444_SDR	-26.71%	-27.45%	-25.25%	-25.64%	-31.20%
	Class ECF2_422_SDR	-23.26%	-24.87%	-22.68%	-23.36%	-28.80%
	Class ECF3_444_HDR	-28.88%	-31.59%	-25.17%	-26.70%	-32.68%
	Class ECF4_422_HDR	-34.59%	-36.31%	-31.78%	-33.12%	-41.75%
	Class ECF5_444_YCgCo	-19.78%	-21.35%	-18.74%	-18.59%	-23.12%
	Class ECF6_444_SCC	-32.10%	-33.11%	-32.25%	-33.37%	-37.44%
	Overall Mandatory 444	-26.34%	-27.76%	-24.30%	-24.94%	-30.47%
	Overall Mandatory 422	-27.04%	-28.69%	-25.71%	-26.61%	-33.12%
	Overall	-27.81%	-29.23%	-26.50%	-27.33%	-32.91%

LD	Class ECF1_444_SDR	-23.66%	-24.36%	-22.39%	-22.13%	-25.38%
	Class ECF2_422_SDR	-19.86%	-21.69%	-19.32%	-19.48%	-22.40%
	Class ECF3_444_HDR	-26.71%	-29.53%	-23.01%	-23.91%	-27.01%
	Class ECF4_422_HDR	-28.76%	-30.56%	-26.02%	-27.45%	-31.20%
	Class ECF5_444_YCgCo	-18.68%	-20.38%	-18.89%	-17.86%	-21.97%
	Class ECF6_444_SCC	-32.01%	-32.96%	-31.26%	-32.35%	-36.08%
	Overall Mandatory 444	-23.82%	-25.27%	-22.07%	-22.02%	-25.35%
	Overall Mandatory 422	-22.83%	-24.65%	-21.55%	-22.14%	-25.33%

SDP=1

	Summary	PSNR		SSIM	MS-SSIM	VMAF
		Y	YUV	Y	Y	Y
Intra	Class ECF1_444_SDR	-19.76%	-19.98%	-19.94%	-19.42%	-22.30%
	Class ECF2_422_SDR	-15.89%	-17.25%	-15.54%	-15.33%	-19.02%
	Class ECF3_444_HDR	-19.32%	-22.43%	-16.46%	-17.61%	-21.10%
	Class ECF4_422_HDR	-22.30%	-23.38%	-19.77%	-21.39%	-24.18%
	Class ECF5_444_YCgCo	-15.40%	-15.74%	-14.56%	-15.21%	-19.99%
	Class ECF6_444_SCC	-35.68%	-36.77%	-32.35%	-35.27%	-36.24%
	Overall Mandatory 444	-19.01%	-20.08%	-18.18%	-18.30%	-21.63%
	Overall Mandatory 422	-18.03%	-19.30%	-16.95%	-17.35%	-20.74%
	Overall	-22.19%	-23.33%	-20.74%	-21.56%	-24.40%

RA	Class ECF1_444_SDR	-26.81%	-27.56%	-25.64%	-25.95%	-31.24%
	Class ECF2_422_SDR	-23.14%	-24.54%	-22.63%	-23.26%	-28.98%
	Class ECF3_444_HDR	-28.78%	-31.23%	-25.05%	-26.58%	-32.85%
	Class ECF4_422_HDR	-34.52%	-36.06%	-31.74%	-33.05%	-41.89%
	Class ECF5_444_YCgCo	-19.56%	-21.11%	-18.63%	-18.54%	-22.72%
	Class ECF6_444_SCC	-35.59%	-36.66%	-35.27%	-36.53%	-38.92%
	Overall Mandatory 444	-26.34%	-27.69%	-24.47%	-25.07%	-30.48%
	Overall Mandatory 422	-26.93%	-28.38%	-25.67%	-26.52%	-33.29%
	Overall	-28.52%	-29.84%	-27.20%	-28.03%	-33.29%

LD	Class ECF1_444_SDR	-23.93%	-24.73%	-22.77%	-22.54%	-25.38%
	Class ECF2_422_SDR	-19.96%	-21.56%	-19.42%	-19.60%	-22.70%
	Class ECF3_444_HDR	-26.46%	-29.06%	-22.72%	-23.62%	-26.35%
	Class ECF4_422_HDR	-28.74%	-30.33%	-26.07%	-27.42%	-31.46%
	Class ECF5_444_YCgCo	-18.37%	-19.95%	-18.82%	-17.79%	-21.78%
	Class ECF6_444_SCC	-35.17%	-36.19%	-34.08%	-35.28%	-37.30%
	Overall Mandatory 444	-23.86%	-25.28%	-22.19%	-22.17%	-25.14%
	Overall Mandatory 422	-22.89%	-24.48%	-21.64%	-22.21%	-25.62%

SDP=0

Preliminary Subjective Test Result

- Subjective Test (Conducted internally in Google) on UHD content:
 - Anchor: (modified) AV1
 - Test: AVM-v10.0.0

Sequences	PSNR BD-rate				VMAF BD-rate	VMAF neg BD-rate	On-site MOS	
	Y	U	V	YUV (14:1:1)			BD-rate	CI
Marathon2	-22.16%	-28.06%	-25.19%	-22.84%	-25.89%	-24.70%	-33.02%	0.54
MountainBay2	-23.03%	-35.88%	-40.07%	-24.93%	-31.67%	-29.97%	-32.08%	0.61
TallBuildings2	-32.88%	-57.88%	-69.84%	-37.06%	-31.27%	-31.47%	-35.82%	0.50
YonseiS01	-22.64%	-24.10%	-28.54%	-22.93%	-28.23%	-26.52%	-31.43%	0.52
meridian	-33.96%	-39.82%	-60.76%	-35.30%	-45.78%	-42.20%	-45.06%	0.49
DrivingPOV3	-34.74%	-37.56%	-60.03%	-35.81%	-42.81%	-41.63%	-50.63%	0.60
Average	-28.24%	-37.22%	-47.40%	-29.81%	-34.27%	-32.75%	-38.01%	0.55

DCR
11-point scale

Takeaways

- Objective coding gain for the latest AV2 release more than 30% (for Random Access) over the AV1
- Comprehensive verification test and subjective evaluation is working in progress.

Reference

- AOM CTC v8.0:
https://groups.aomedia.org/g/wg-codec/files/InputDocuments/F2025/CWG-F384/CWG-F384_AV2_CTC_v8.pdf
- Encoder and Decoder code base:
 - AOM Video Model: <https://gitlab.com/AOMediaCodec/avm>. Latest release is [AVM v12.0 anchor](#)
- HDRTools is used for downscaling and upscaling in adaptive streaming configuration.
 - <https://gitlab.com/standards/HDRTools/-/tree/v0.22>
- VMAF [v2.3.1](#) is used for quality metrics calculation.
- High level python script for CTC Test is available under
https://gitlab.com/AOMediaCodec/avm/-/tree/main/tools/convexhull_framework
- Excel spreadsheet template for presenting and reporting the CTC result is available under
https://gitlab.com/AOMediaCodec/avm/-/tree/main/tools/convexhull_framework/bin
- CTC results are available under
https://gitlab.com/AOMediaCodec/avm/-/tree/main/tools/convexhull_framework/ctc_result?ref_type=heads



Thank You