Deep Video Precoding: Toward A Generalized Deep Perceptual Optimizer

Technical presentation: AOMedia Research Symposium

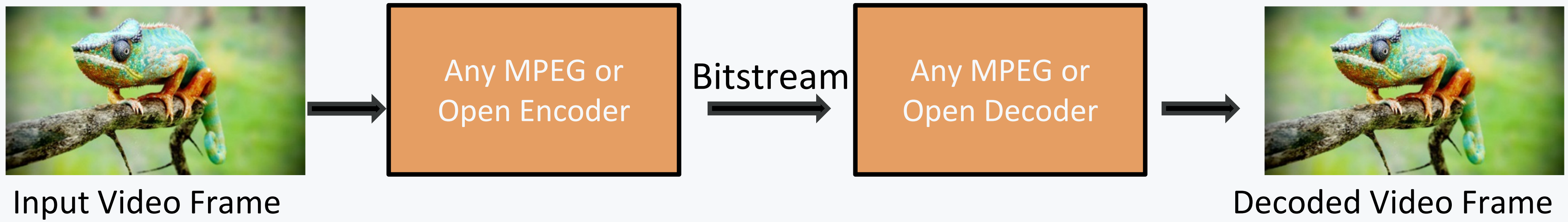
Aaron Chadha, Russell Anam, Eirina Bourtsoulatze, Ilya Fadeev, Vasileios Giotsas, Yiannis Andreopoulos

October 2019

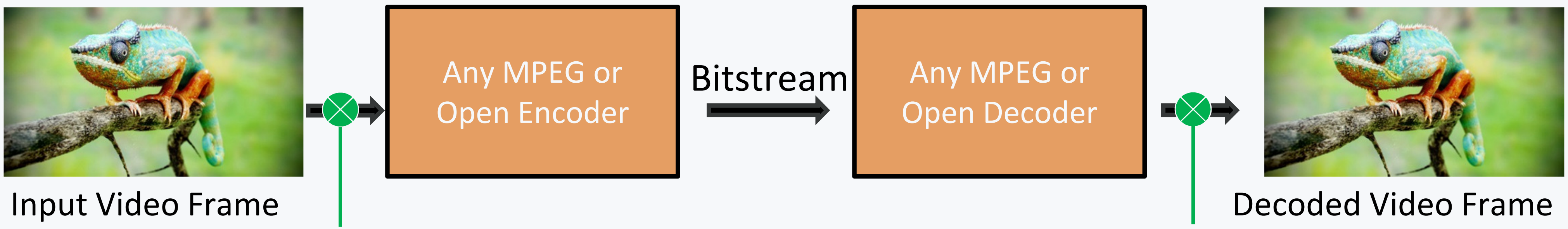
SUMMARY OF WHAT WE DO

Our solution can be applied either on both sides, or on the encoding side only. For more technical details please visit our website
https://www.isize.co/bitsave/
https://www.isize.co/upscale/

Conventional Codec Pipeline: Full resolution H.264/AVC – HEVC – VP9 – AV1 – VVC codec



iSIZE + Conventional Codec Pipeline: Deep Video Precoding and Post-Decoding Enhancement

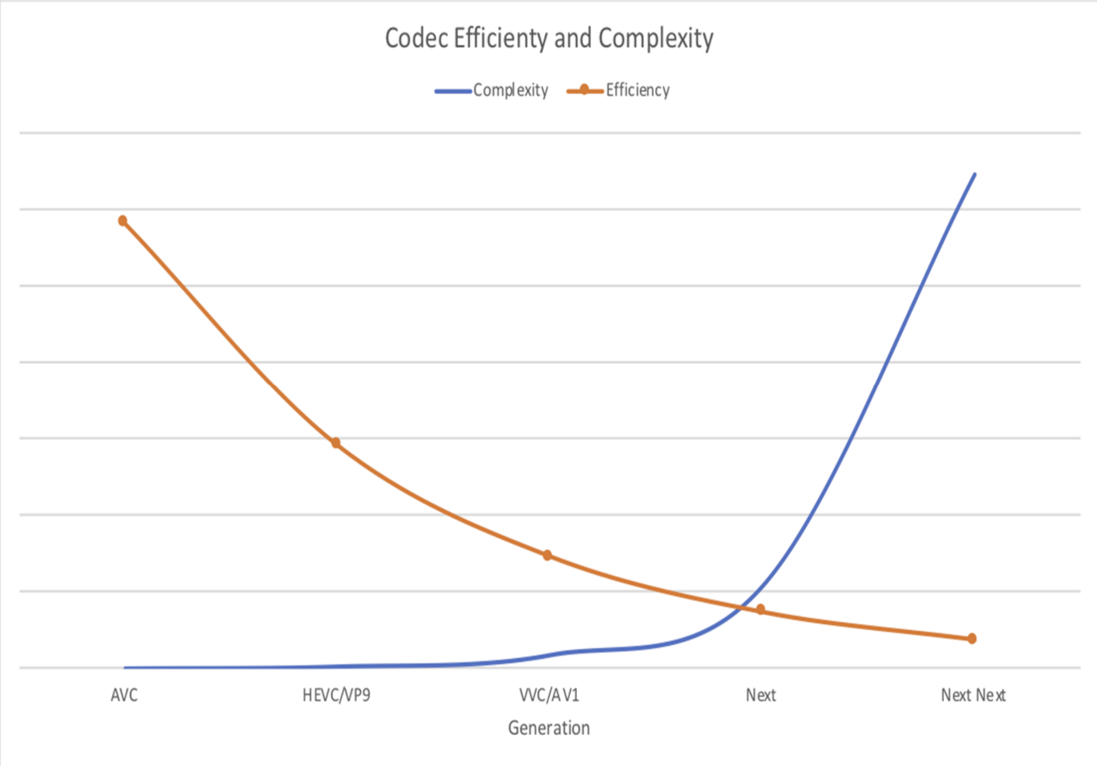


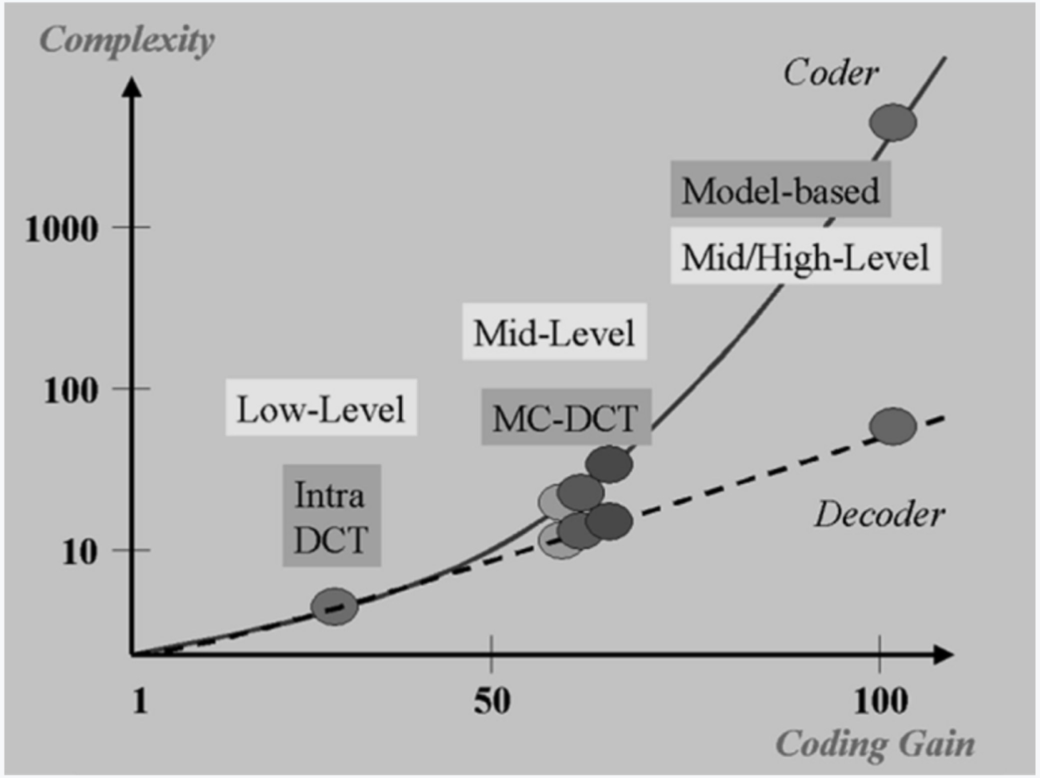
iSIZE Precoder:
Deep Perceptual Optimizer

Perceptual Quality Metrics
(VMAF, VIF, DLM,MS-SSIM…),
included in VQEG and
used by Netflix, video standardization, etc.

iSIZE Post-Decoder
(OPTIONAL)

SUMMARY OF WHY WE DO IT





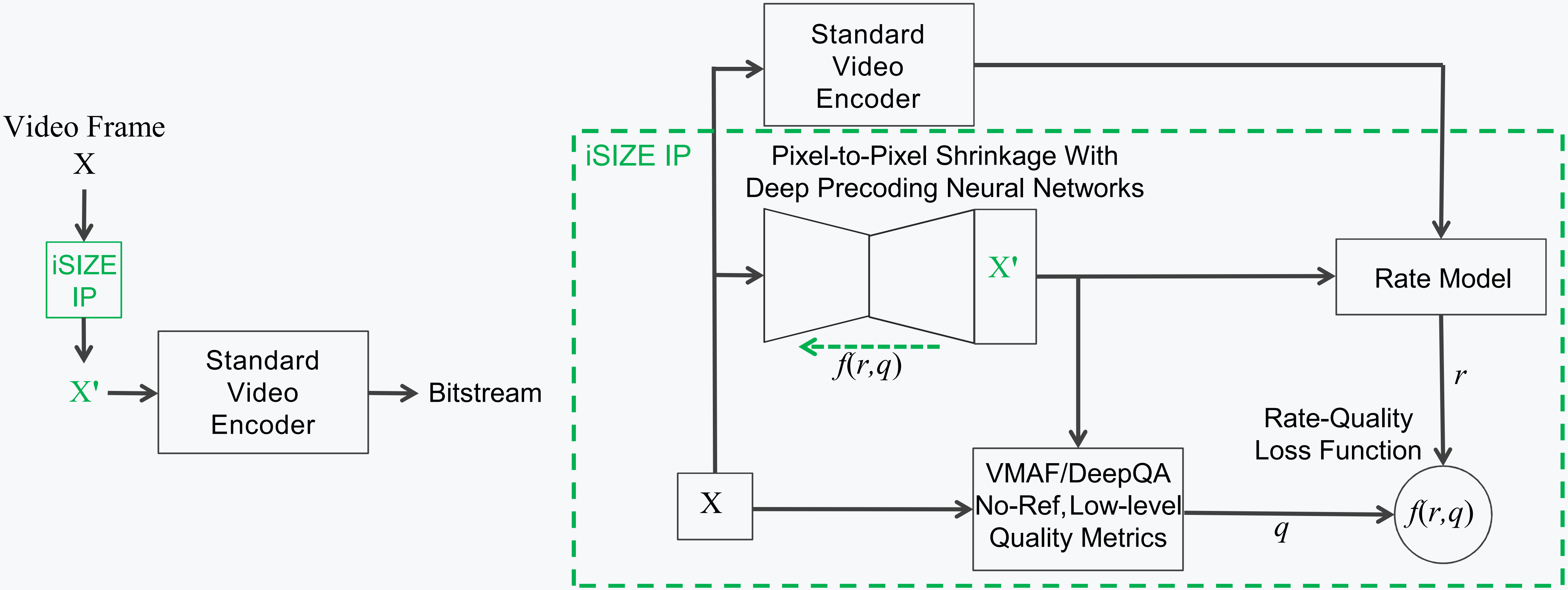
https://www.linkedin.com/pulse/encoder-complexity-hits-wall-david-ronca/ (D. Ronca, Facebook, 2019)

Sikora, Proc. of the IEEE, 2005, https://doi.org/10.1109/JPROC.2004.839601

- Moore's law (mainly device power+heat dissipation) and cloud-based scaling **have both hit the wall**
- **We are at an inflection point:** perceptual metrics and DL are mature enough to allow for robust perceptual precoding
- **Codecs are amazing SNR/SSIM-to-bitrate machines, but these loss functions have significant limitations**

OUR GOAL: REVERSE-ENGINEER HUMAN VISUAL PERCEPTION

- Deep video precoding is a **deep perceptual optimizer using neural networks**:

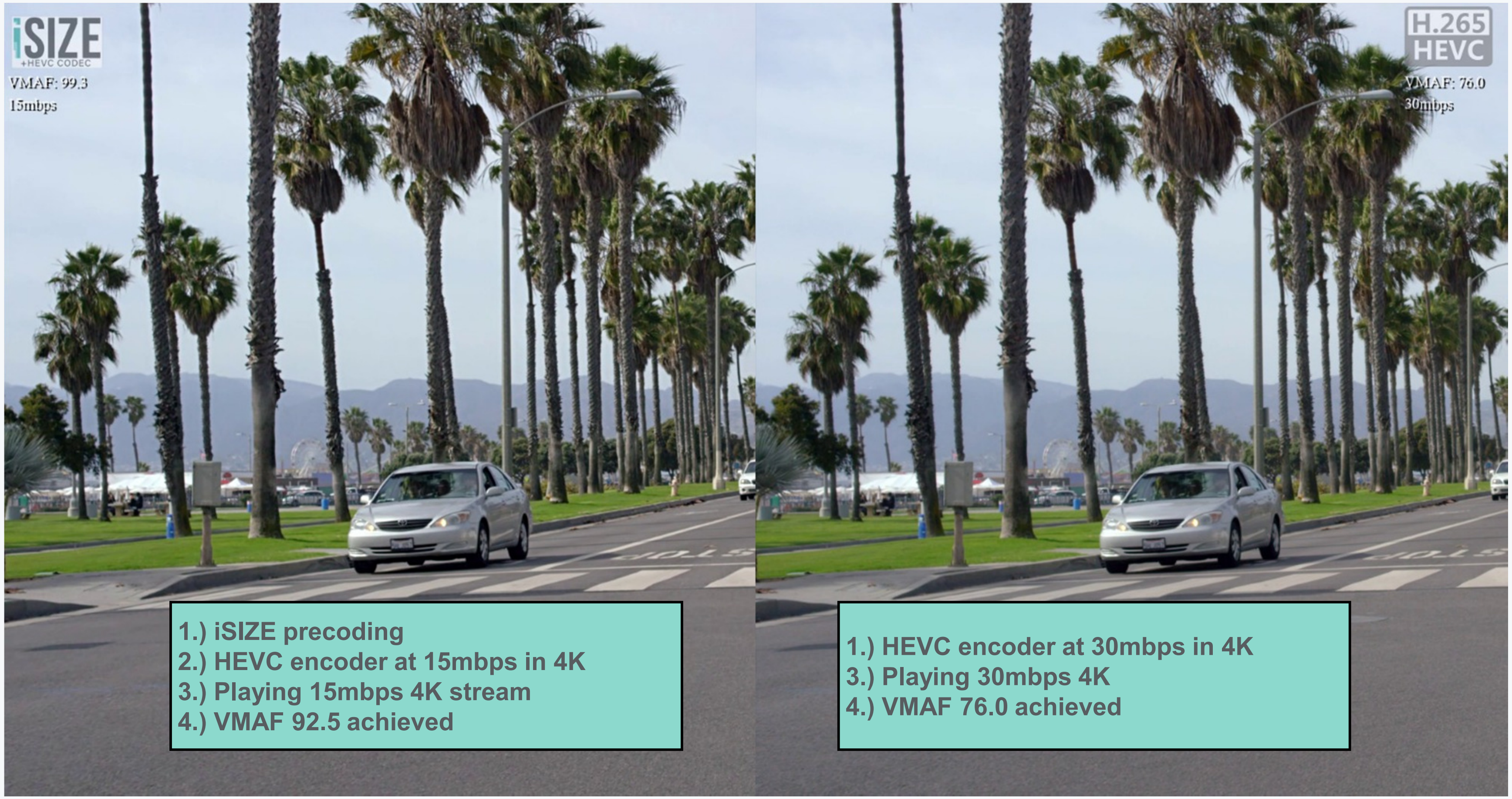


- This escapes the legacy constraints of video encoding standards and is ***codec independent***
- It enables AI to penetrate video encoding & delivery with backward compatibility and ***single-pass stream processing***
- It creates an ***adversarial approach*** between deep video precoding and advanced perceptual metrics

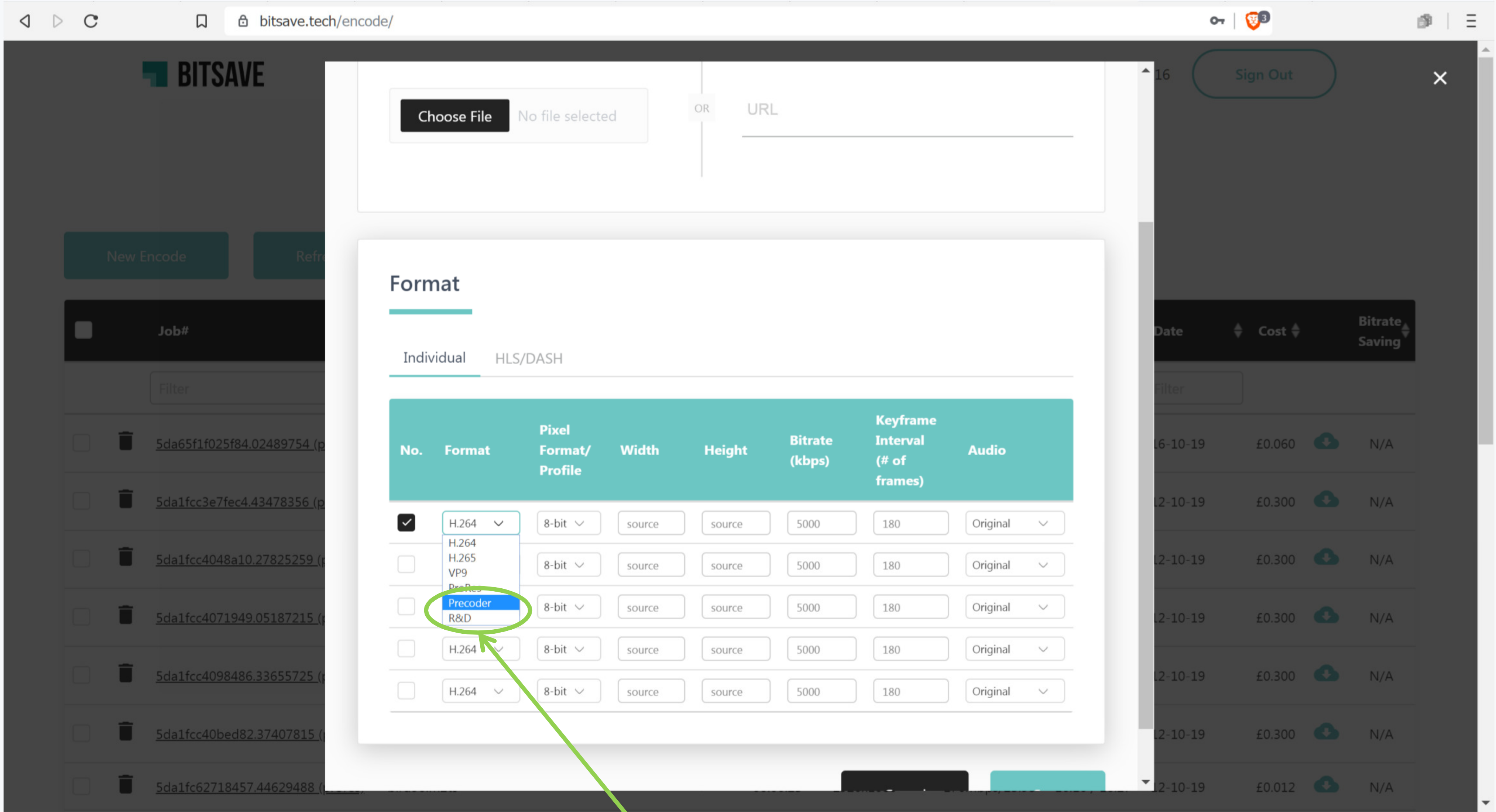
→ This may allow us to completely ***reverse-engineer human visual perception of VoD and livestreamed content***

ISIZE PRECODER – AVAILABLE AT BITSAVE.TECH

- Our demo shows content that has been **precoded by iSize Precoding** and encoded by HEVC in 4K resolution at half the bitrate without any loss of VMAF compared with normal 4K HEVC encoded content



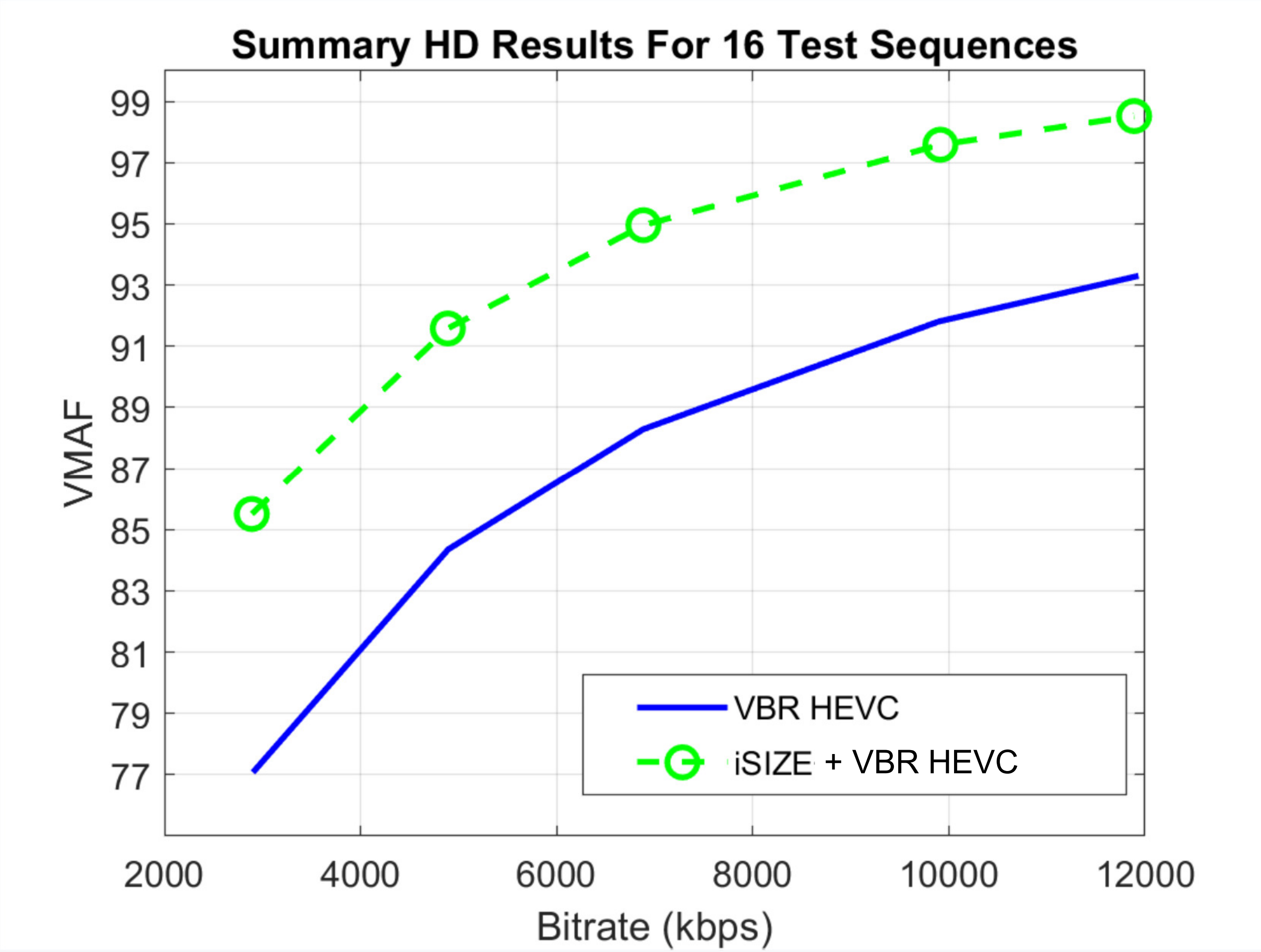
ISIZE PRECODER – FREE TRIAL AVAILABLE AT BITSAVE.TECH



Select “Precoder” or “R&D” to get the precoder output and use your own encoder to compare against your own recipe

RESULTS ON FHD CONTENT

HEVC VBR, TuneVMAF, HD resolution, 3Mbps-12Mbps



Notes:

1. For fair comparison, we use the same codec settings.
2. VMAF is a state-of-the-art perceptual quality metric proposed by Netflix and recognized by VQEG, see:

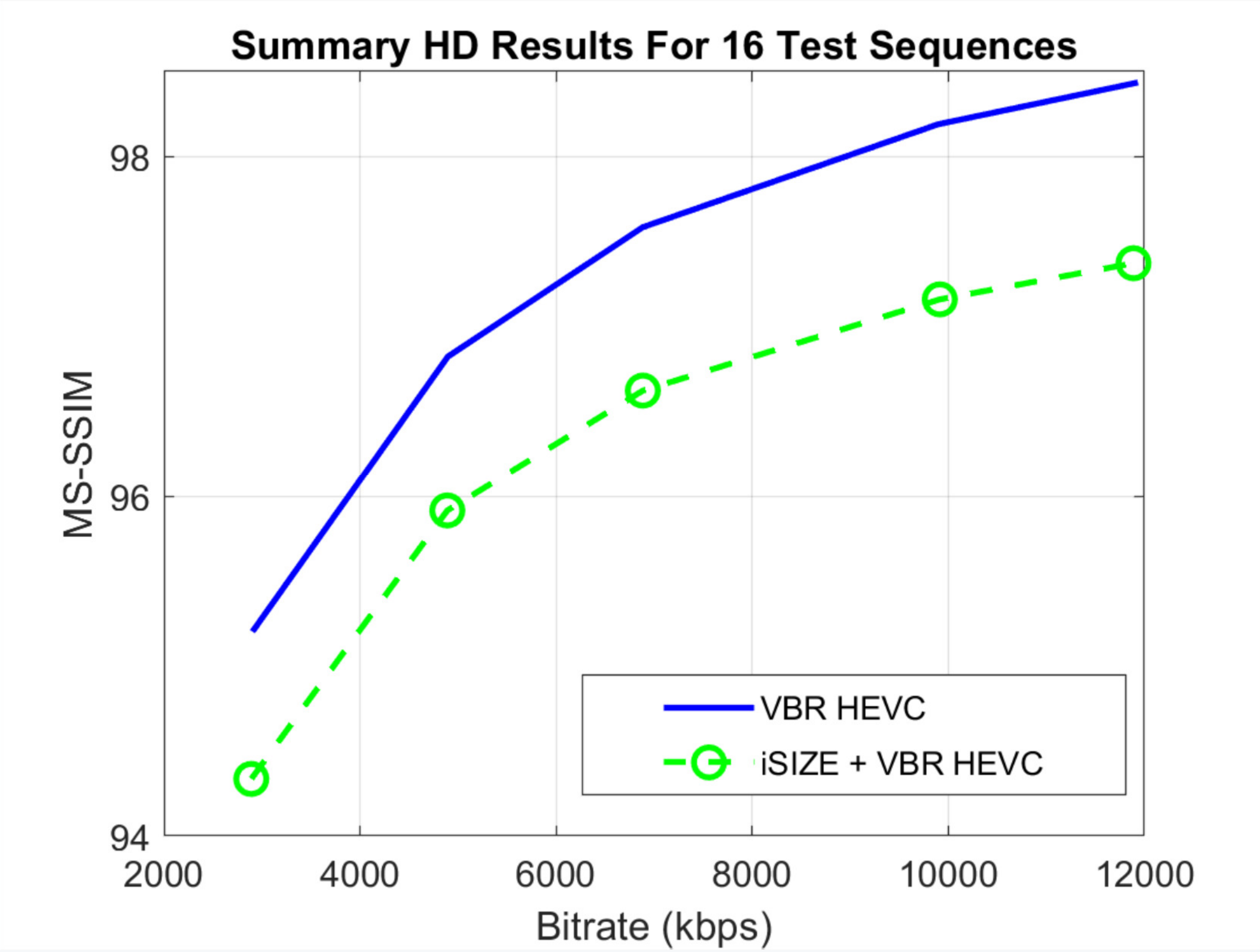
https://vqeg.github.io/software-tools/quality%20analysis/ffmpeg-quality-metrics/

https://medium.com/netflix-techblog/vmaf-the-journey-continues-44b51ee9ed12

3. **BD-rate is -47% or 6.9 VMAF points.**
4. VMAF of 6 points is the JND threshold
5. We use arithmetic-mean VMAF (FFmpeg libvmaf model v0.6.1) but similar or higher ΔVMAF was obtained with harmonic mean

RESULTS ON FHD CONTENT: IN MORE DETAIL

HEVC VBR, TuneVMAF, HD resolution, 3Mbps-12Mbps

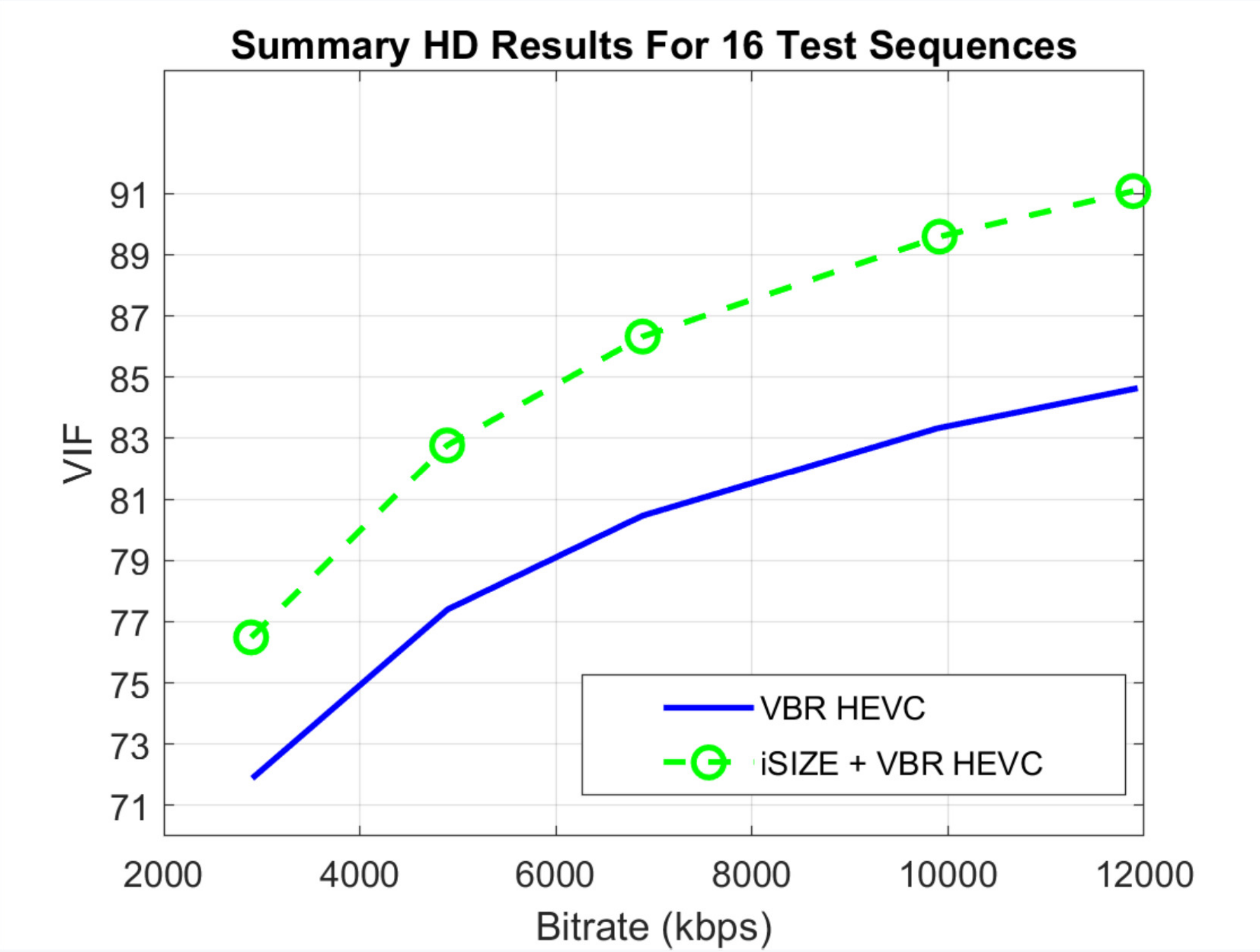


Notes:

1. For fair comparison, we use the same codec settings.
2. MS-SSIM is the multi-scale structural similarity index metric.
3. **BD-rate is 47% or 0.9% of MS-SSIM.**

RESULTS ON FHD CONTENT: IN MORE DETAIL

HEVC VBR, TuneVMAF, HD resolution, 3Mbps-12Mbps

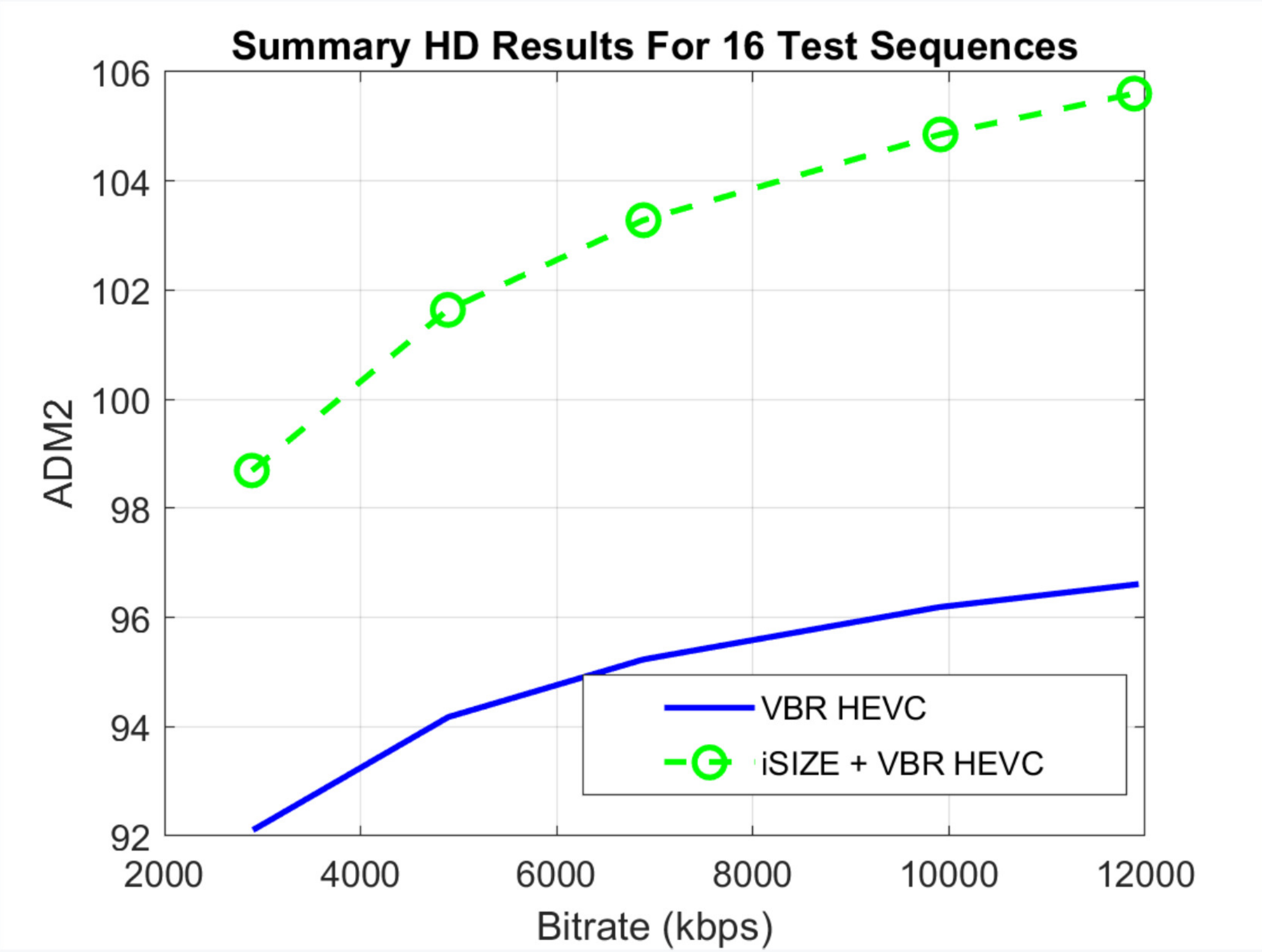


Notes:

1. For fair comparison, we use the same codec settings.
2. VIF is the visual information fidelity metric.
3. **BD-rate is -39% or 5.3 points.**

RESULTS ON FHD CONTENT: IN MORE DETAIL

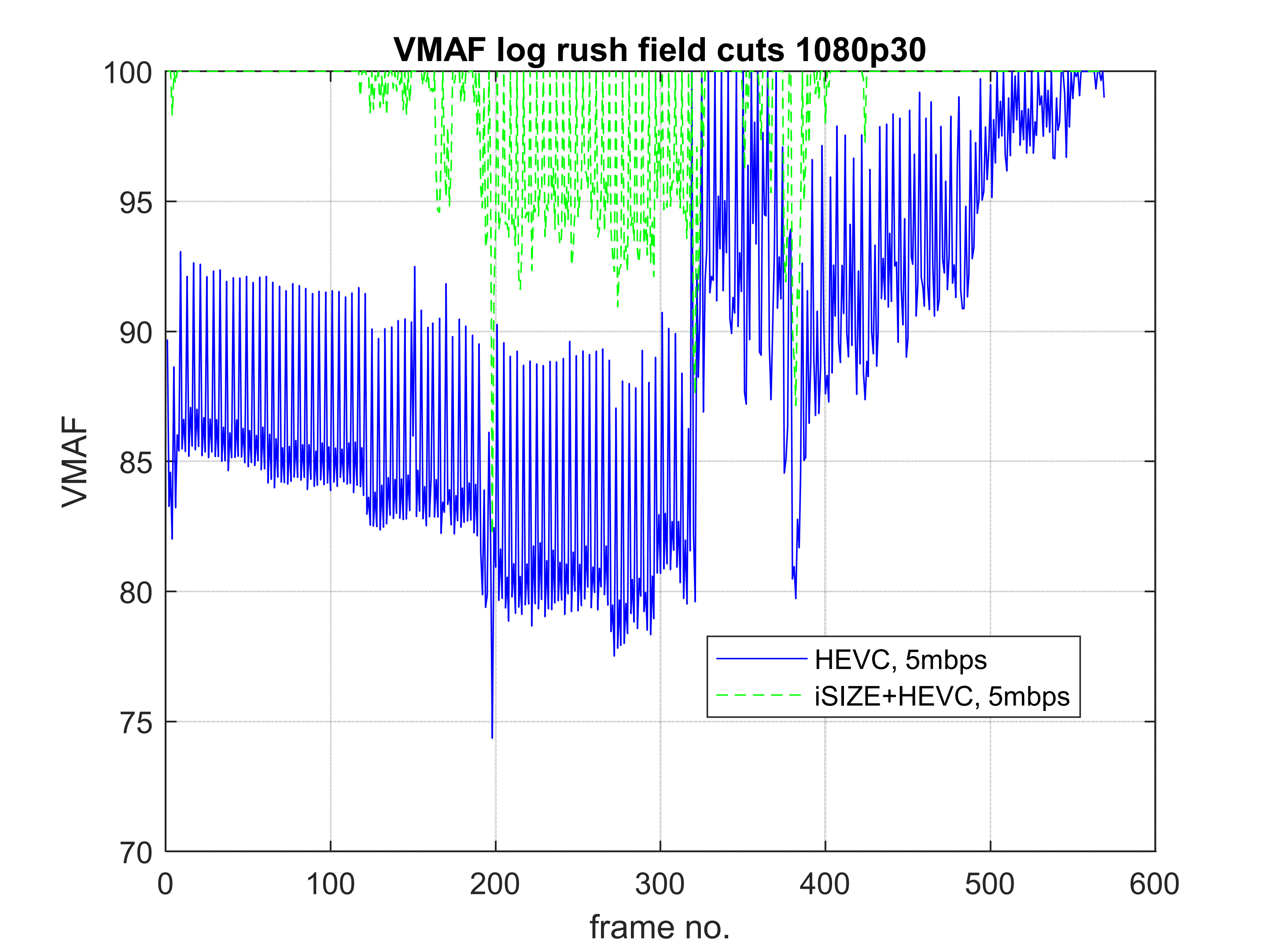
HEVC VBR, TuneVMAF, HD resolution, 3Mbps-12Mbps



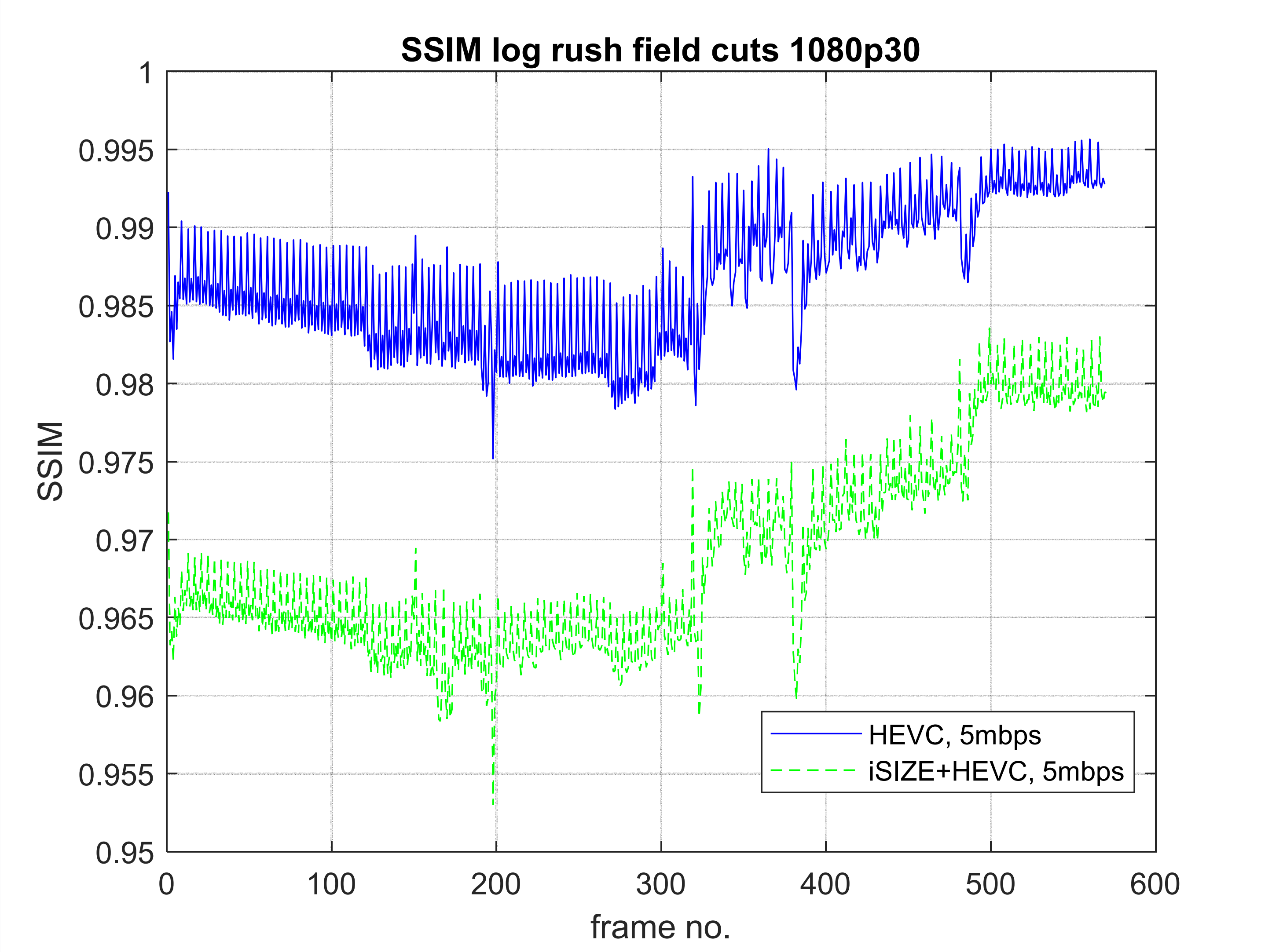
Notes:

1. For fair comparison, we use the same codec settings.
2. ADM2 is a variant of the DLM (detail loss metric).
3. **BD-rate is -88% or 7.8 points.**

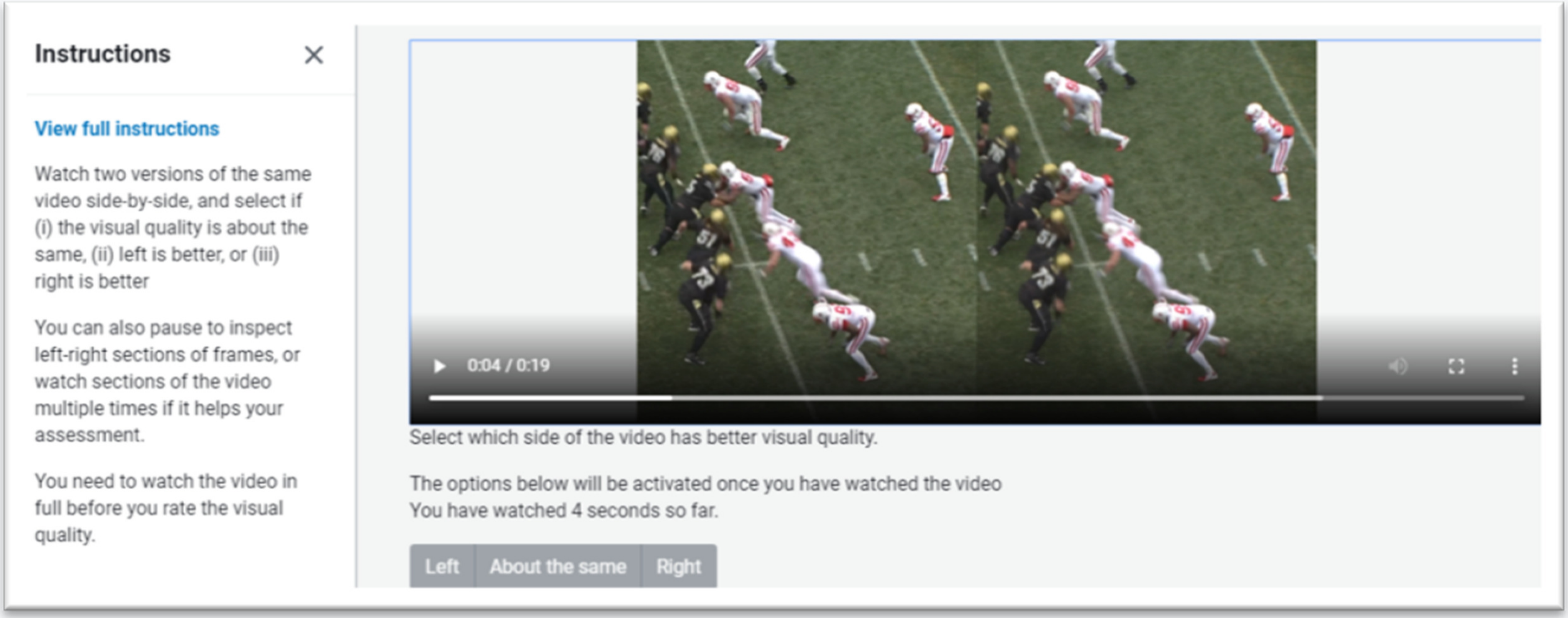
FRAME-BY-FRAME COMPARISON



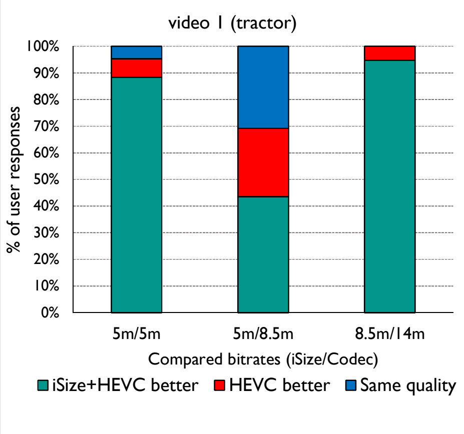
FRAME-BY-FRAME COMPARISON

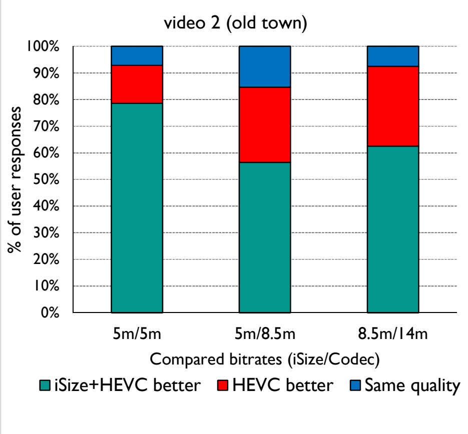


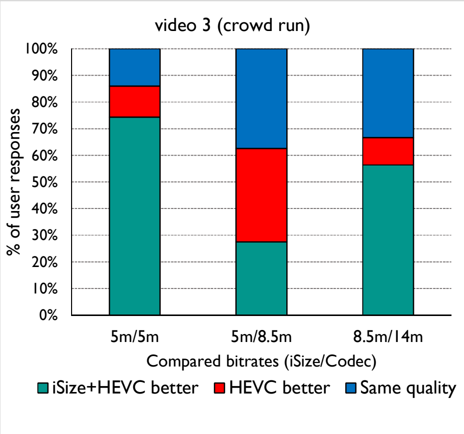
HEVC RESULTS ON FHD CONTENT VIA AMAZON MTURK

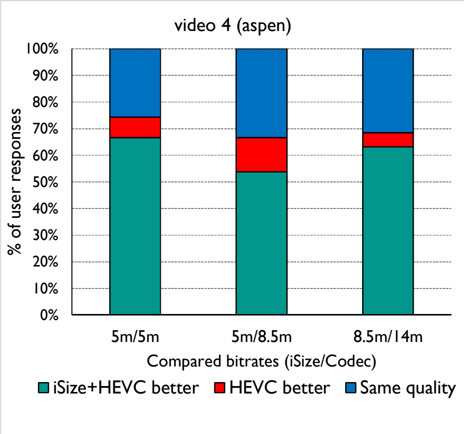


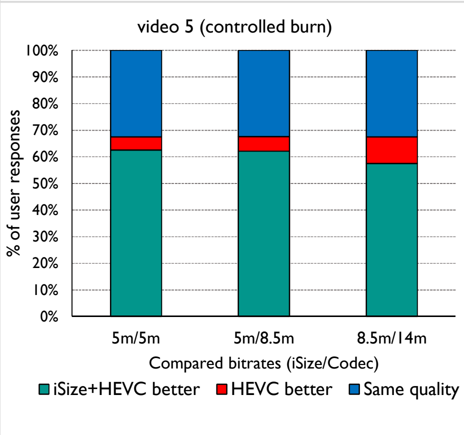
HEVC RESULTS ON FHD CONTENT VIA MTURK

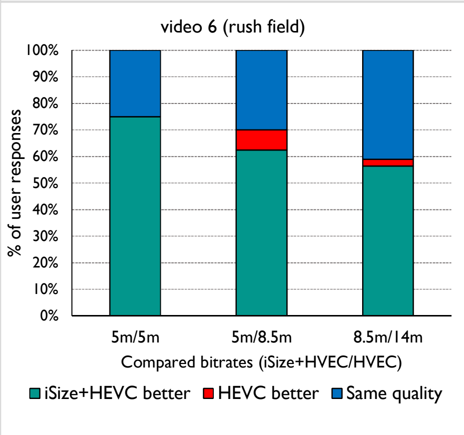










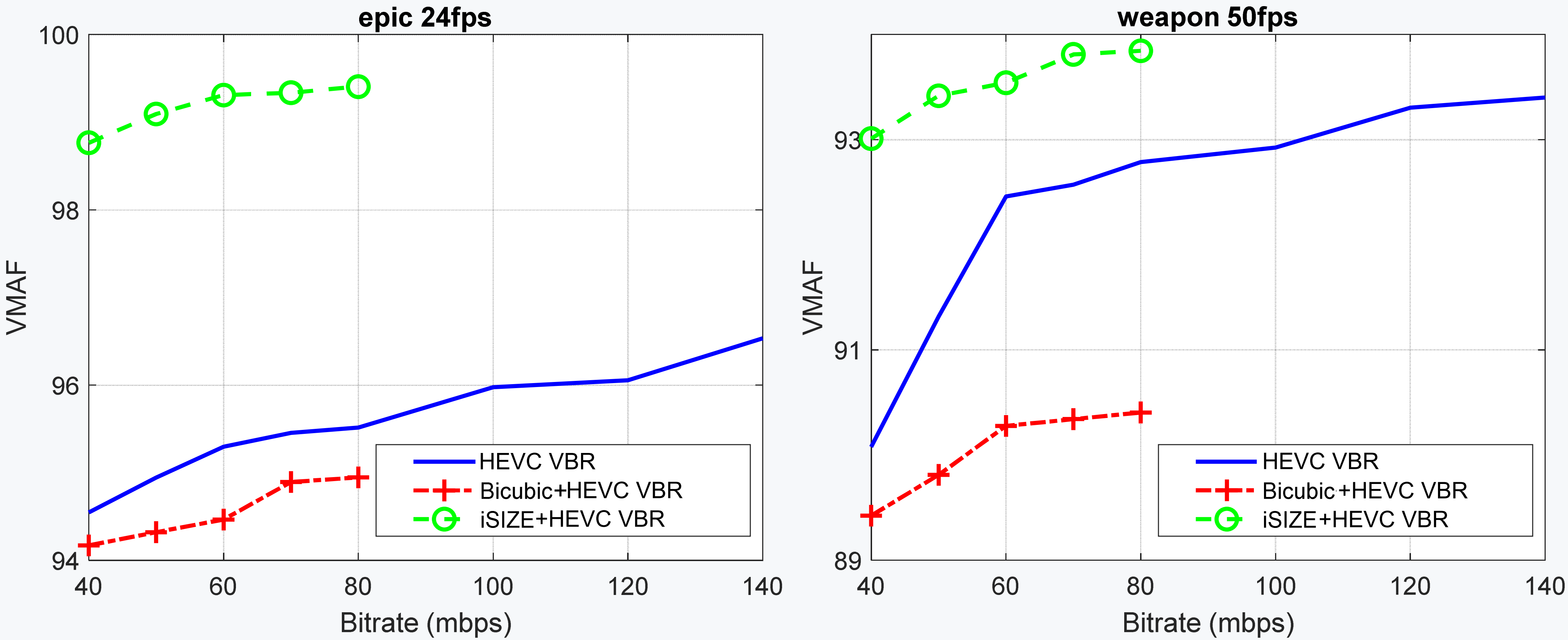


Video Title	Bitrate (iSize/Codec)	iSize+ HEVC Better (%)	HEVC Better (%)	No Preference (%)	ΔVMAF
tractor	5m/5m	88	07	05	0.7
	5m/8.5m	44	26	31	0.3
	8.5m/14m	95	05	00	0.2
old town	5m/5m	79	14	07	7.5
	5m/8.5m	56	28	15	6.4
	8.5m/14m	63	30	08	5.8
crowd run	5m/5m	74	12	14	8.7
	5m/8.5m	28	35	38	-3.3
	8.5m/14m	56	10	33	0.2
aspen	5m/5m	67	08	26	8.6
	5m/8.5m	54	13	33	3.5
	8.5m/14m	63	05	32	4.6
controlled burn	5m/5m	63	05	33	8.2
	5m/8.5m	62	05	32	5.3
	8.5m/14m	58	10	33	3.9
rush field	5m/5m	75	00	25	9.8
	5m/8.5m	63	08	30	4.5
	8.5m/14m	56	03	41	3.4
touchdown pass	5m/5m	46	26	28	4.0
	5m/8.5m	60	28	12	2.6
	8.5m/14m	62	13	30	1.9
Average:	**5m/5m**	**70.2**	**10.2**	**19.6**	**6.8**
	5m/8.5m	**52.3**	**20.4**	**27.3**	**2.8**
	8.5m/14m	**64.4**	**10.6**	**25.0**	**2.9**

A|B videos and full bitstreams:
https://www.isize.co/portfolio/demo

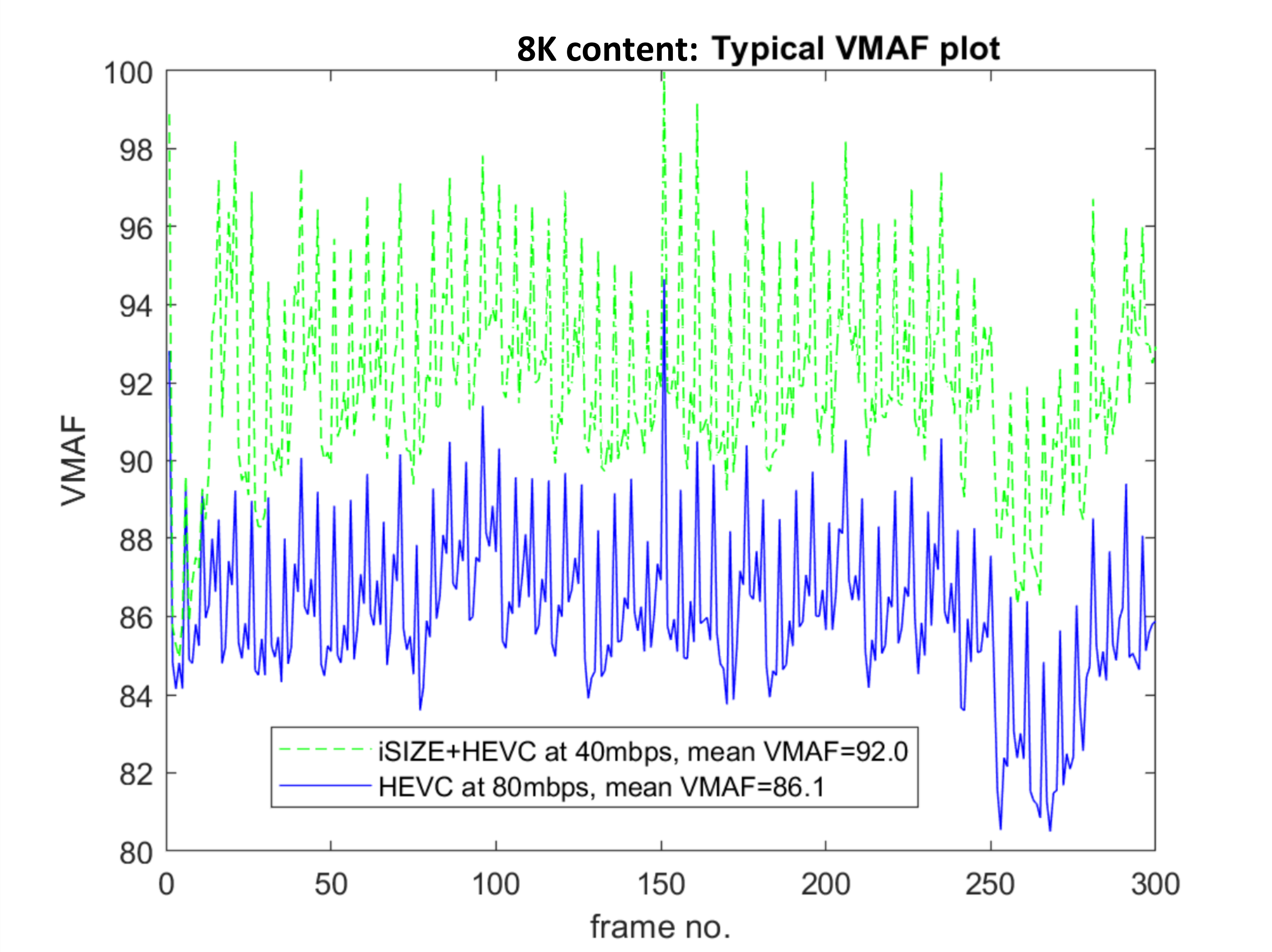
RESULTS ON 8K CONTENT

HEVC VBR, TuneVMAF, 8K resolution

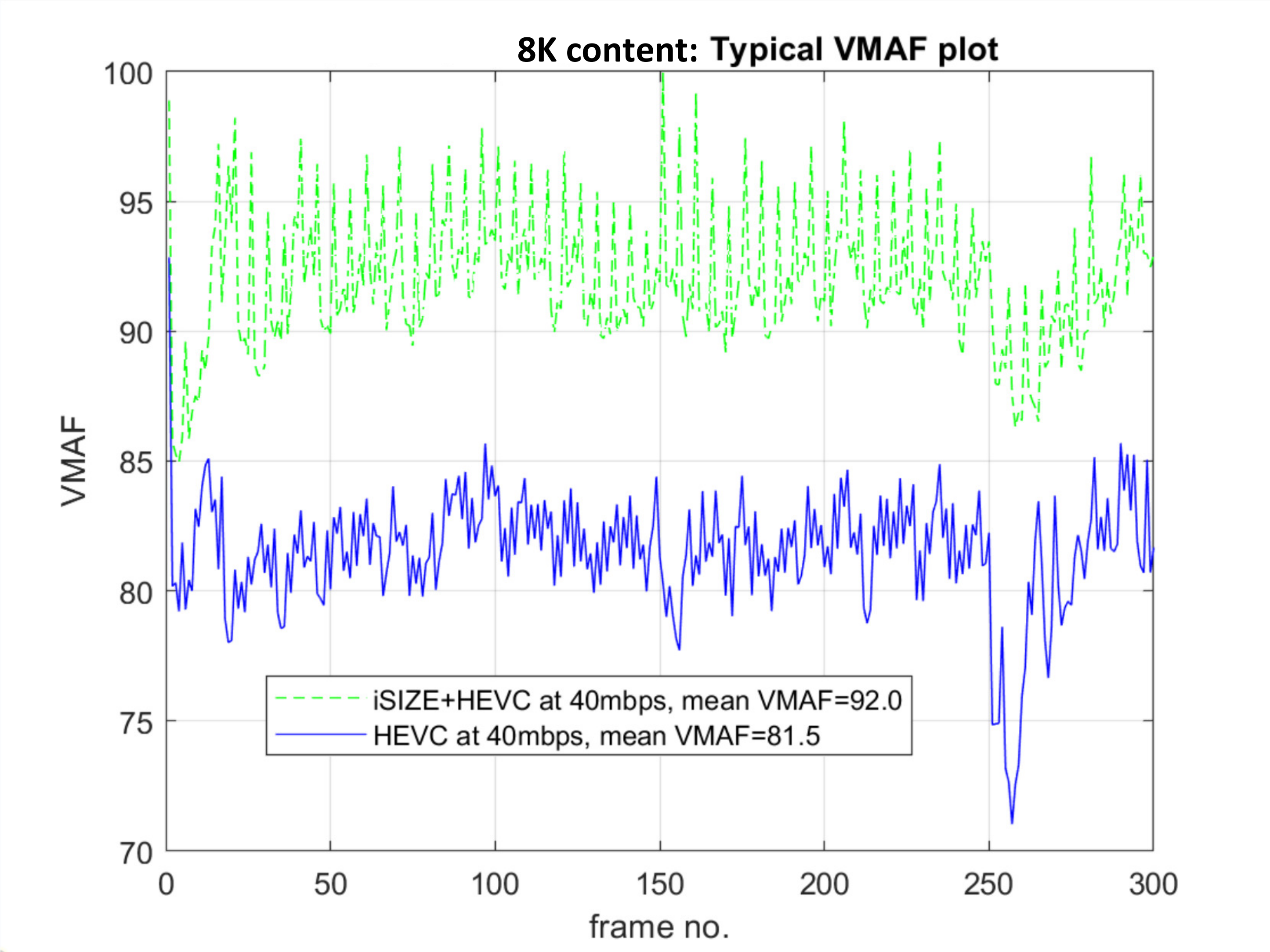


- Clips from red.com
- Downscaling was by factor (4/3,4/3)

FRAME-BY-FRAME COMPARISON

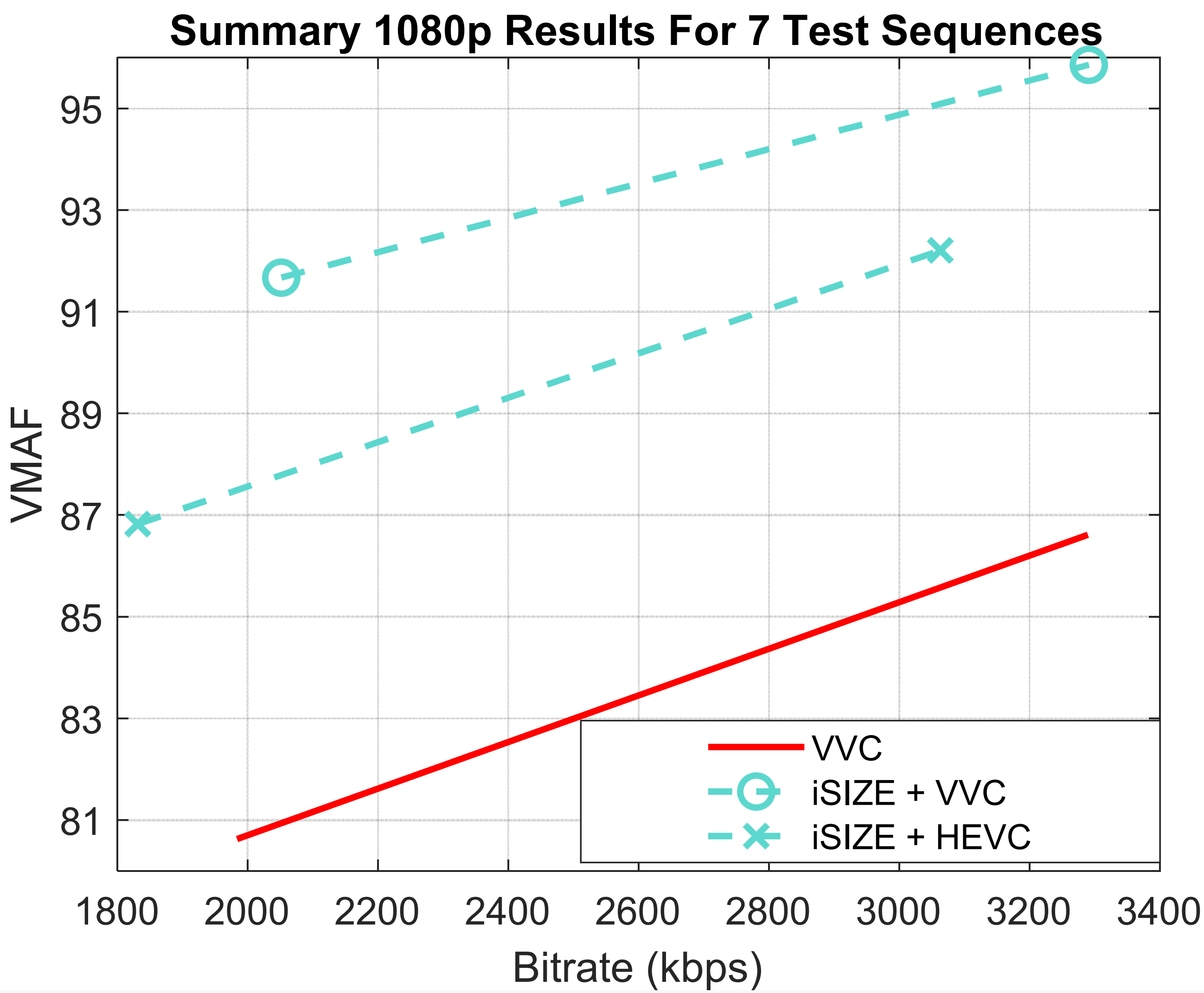


FRAME-BY-FRAME COMPARISON



RESULTS ON FHD CONTENT AND VERSATILE VIDEO CODING

VVC JVET VTM v.6.2rc1, IntraPeriod=64, RateControl=1



Notes:

1. For fair comparison, we use the same codec settings.
2. VMAF is a state-of-the-art perceptual quality metric proposed by Netflix and recognized by VQEG, see:

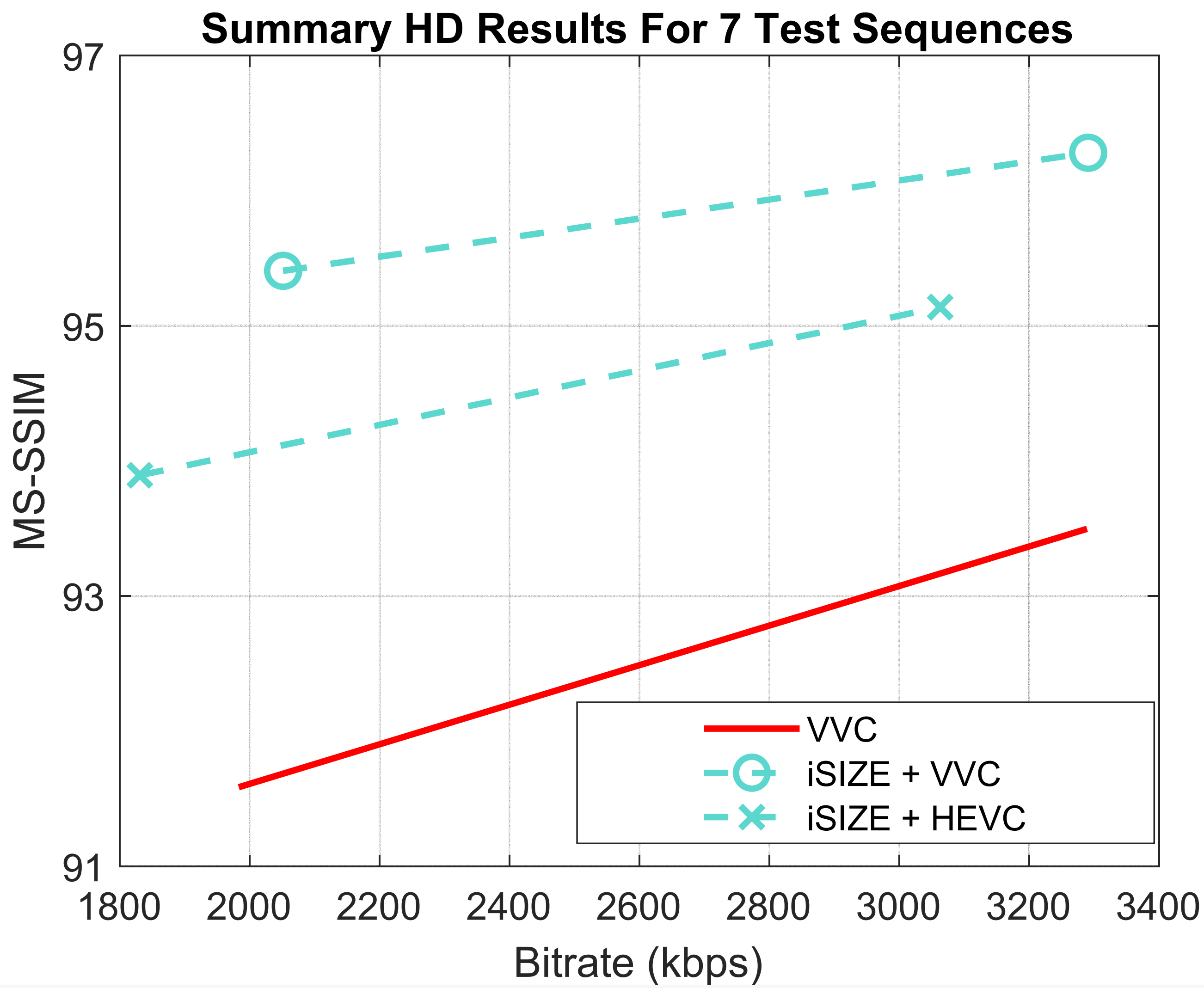
https://vqeg.github.io/software-tools/quality%20analysis/ffmpeg-quality-metrics/

https://medium.com/netflix-techblog/vmaf-the-journey-continues-44b51ee9ed12

3. We use arithmetic-mean VMAF (FFmpeg libvmaf model v0.6.1) but similar or higher ΔVMAF was obtained with harmonic mean

RESULTS ON FHD CONTENT AND VVC: IN MORE DETAIL

VVC JVET VTM v.6.2rc1, IntraPeriod=64, RateControl=1

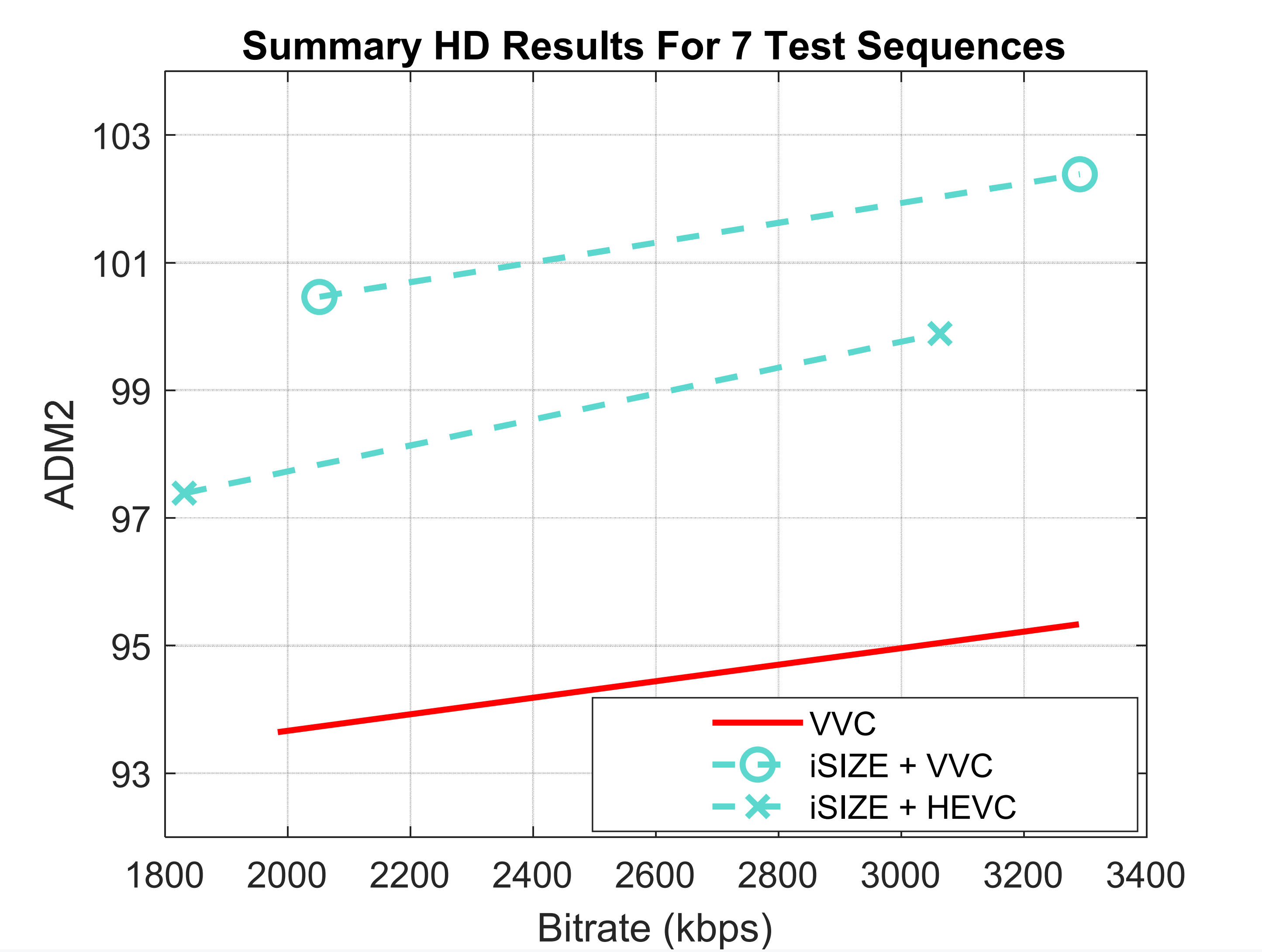


Notes:

1. For fair comparison, we use the same codec settings.
2. MS-SSIM is the multi-scale structural similarity index metric.

RESULTS ON FHD CONTENT AND VVC: IN MORE DETAIL

VVC JVET VTM v.6.2rc1, IntraPeriod=64, RateControl=1

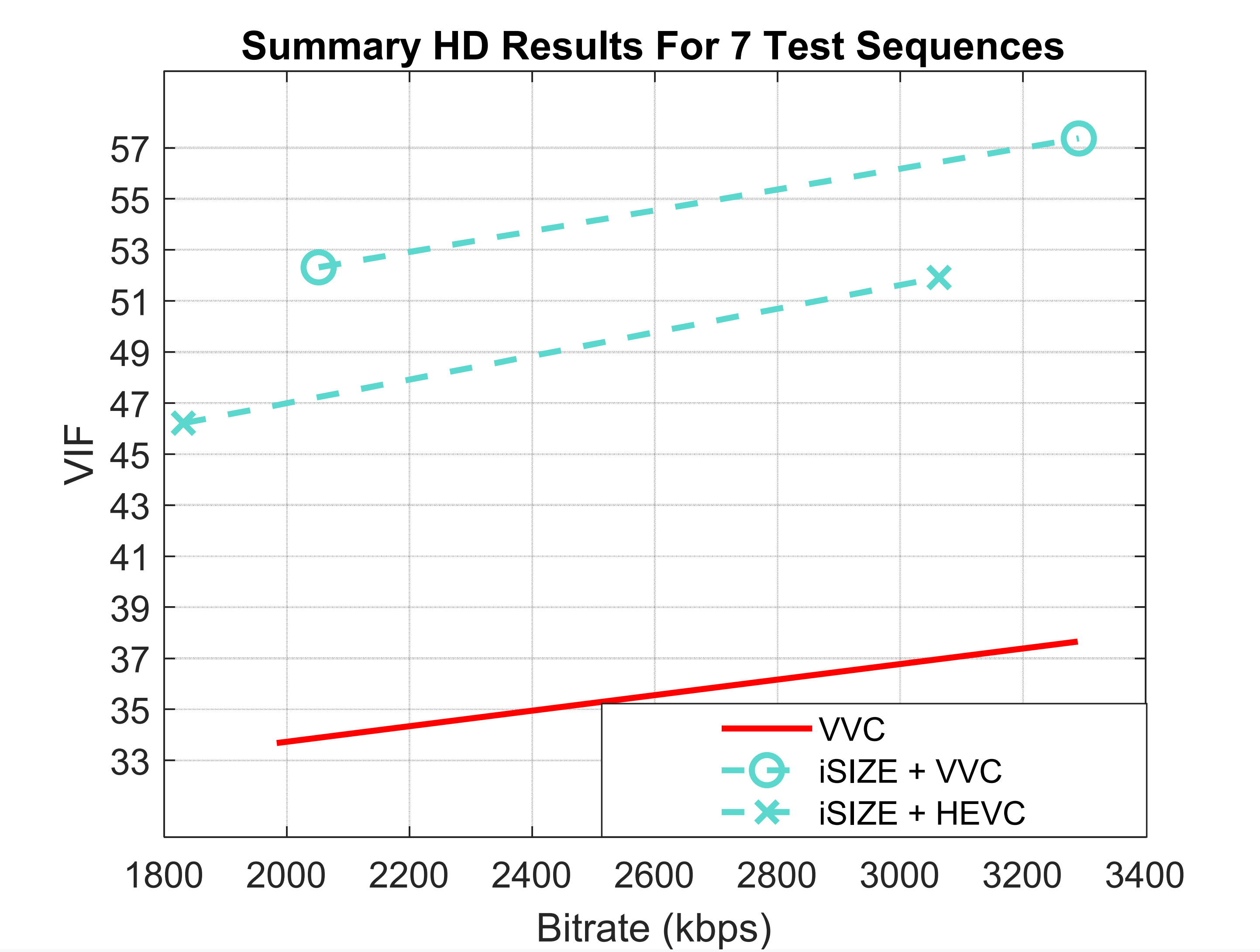


Notes:

1. For fair comparison, we use the same codec settings.
2. VIF is the visual information fidelity metric.

RESULTS ON FHD CONTENT AND VVC: IN MORE DETAIL

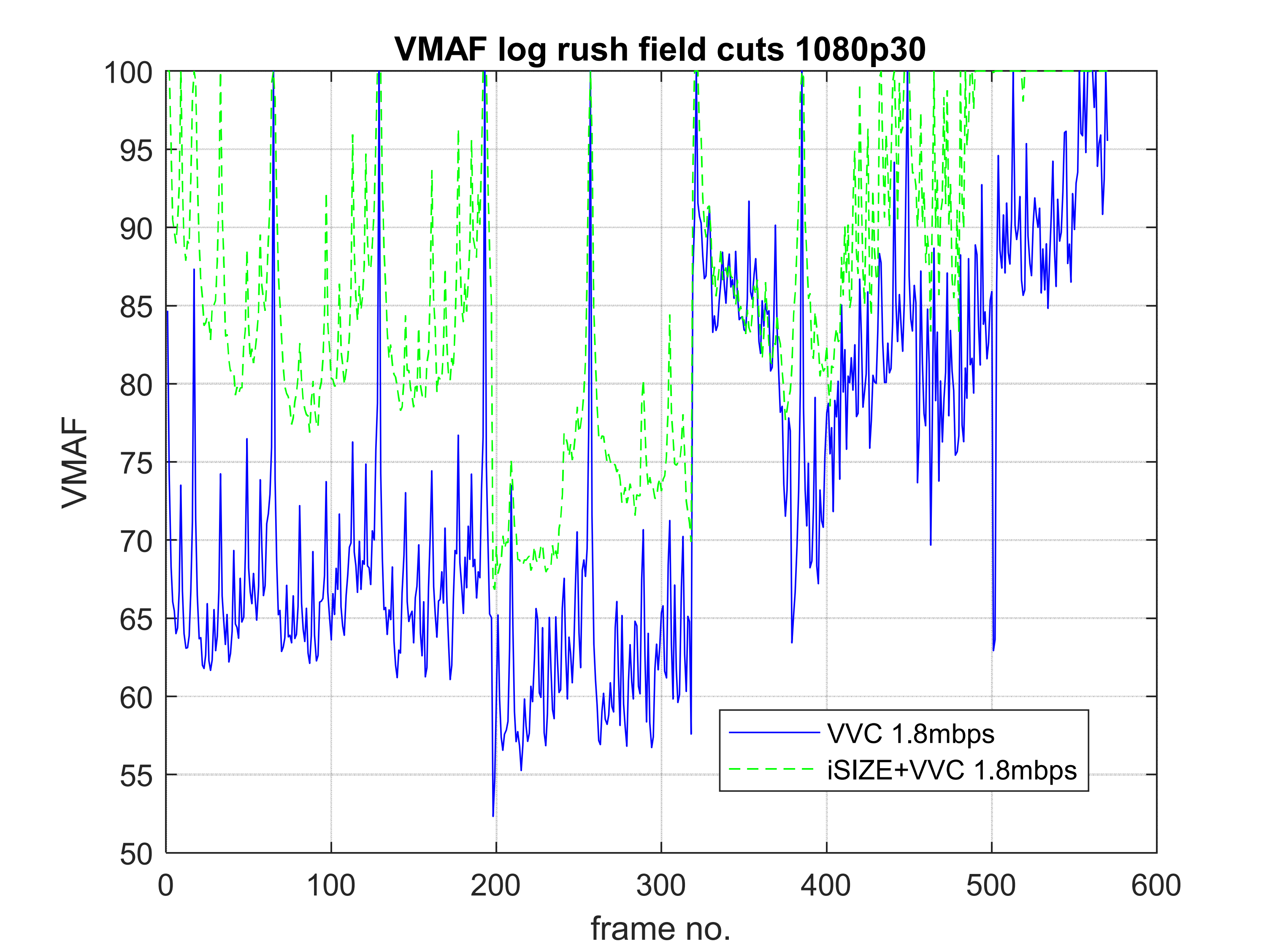
VVC JVET VTM v.6.2rc1, IntraPeriod=64, RateControl=1



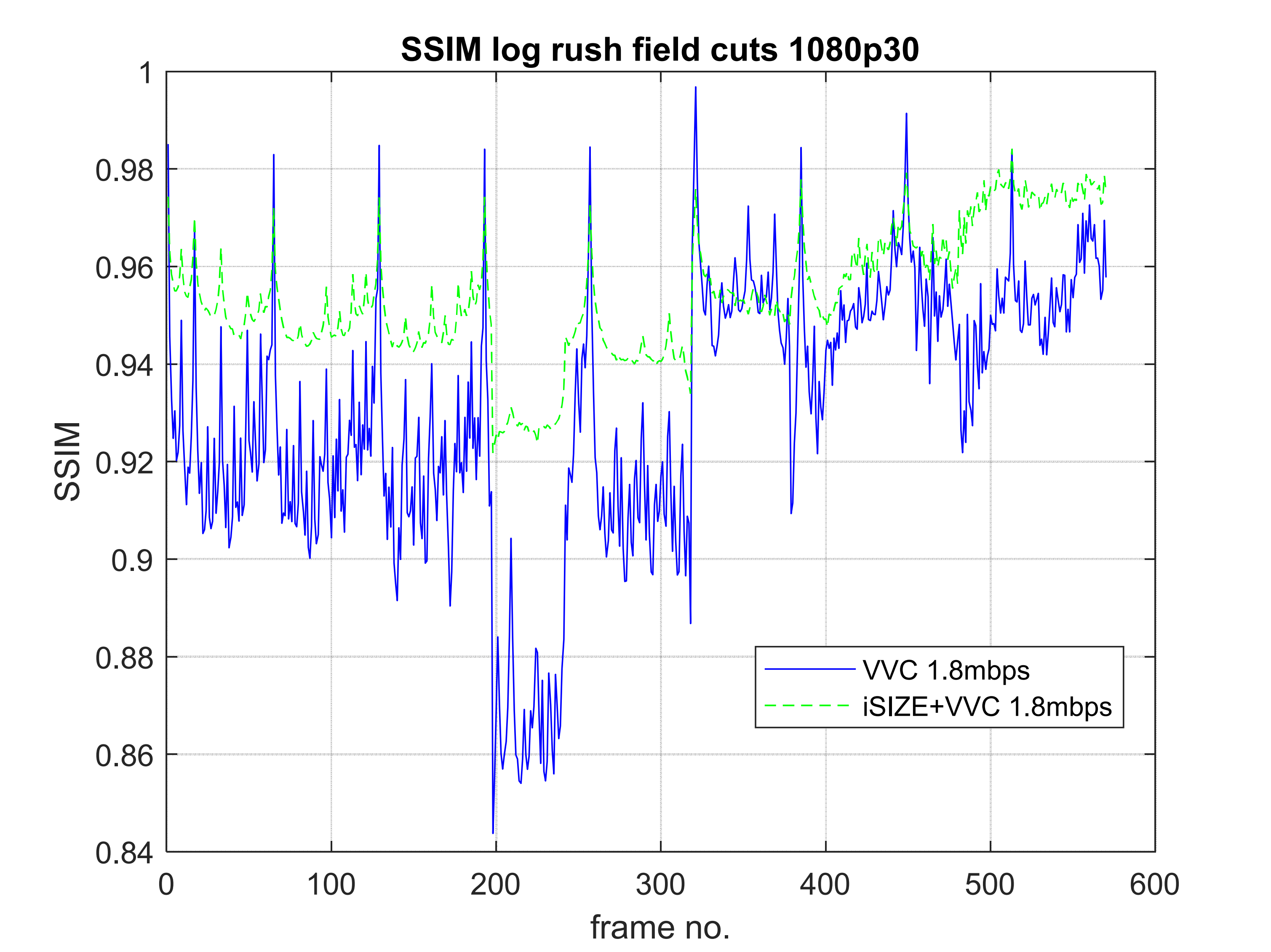
Notes:

1. For fair comparison, we use the same codec settings.
2. ADM2 is a variant of the DLM (detail loss metric).

FRAME-BY-FRAME COMPARISON



FRAME-BY-FRAME COMPARISON



VVC RESULTS ON FHD CONTENT: VIA MTURK

A|B videos and full bitstreams:
https://www.isize.co/portfolio/demo2

Video Title	Bitrate (iSize/VVC)	iSize+Codec Better (%)	VVC Better (%)	No Preference (%)	ΔVMAF
tractor	1.8m/1.8m	73	10	18	1.7
	1.8m/3.0m	65	16	19	0.3
	iSIZE+HEVC 3.0m/ VVC 3.0m	70	14	16	0.3
old town	1.8m/1.8m	70	19	11	12.8
	1.8m/3.0m	65	24	11	9.6
	iSIZE+HEVC 3.0m/ VVC 3.0m	69	21	10	9.6
crowd run	1.8m/1.8m	56	18	26	4.8
	1.8m/3.0m	34	46	20	-4.2
	iSIZE+HEVC 3.0m/ VVC 3.0m	21	71	08	-11.4
aspen	1.8m/1.8m	61	11	28	12.5
	1.8m/3.0m	59	24	18	4.6
	iSIZE+HEVC 3.0m/ VVC 3.0m	56	16	28	11.7
controlled burn	1.8m/1.8m	65	14	21	21.8
	1.8m/3.0m	58	20	23	15.2
	iSIZE+HEVC 3.0m/ VVC 3.0m	59	18	24	15.5
rush field	1.8m/1.8m	71	06	23	12.3
	1.8m/3.0m	68	14	19	2.7
	iSIZE+HEVC 3.0m/ VVC 3.0m	61	19	20	9.3
touchdown pass	1.8m/1.8m	70	16	14	8.4
	1.8m/3.0m	60	23	18	4.2
	iSIZE+HEVC 3.0m/ VVC 3.0m	70	20	10	4.2
Average:	**1.8m/1.8m**	**67**	**13**	**20**	**10.6**
	1.8m/3.0m	**58**	**24**	**18**	**4.6**
	iSIZE+HEVC 3.0m/ VVC 3.0m	**58**	**26**	**16**	**5.6**

FAST IMPLEMENTATION

- 4Core/4Thread 8Gb Ubuntu 18.04 VM (Intel Xeon CPU)
- Under downscaling, we encode/decode a lower-res (downscaled) video.

  - **Runtime and MAC (both per frame) for perceptual precoding for 1080p input resolution:**

Resolution	Time (ms)
3840x2160	284.07
3072x1728	256.74
1920x1080	175.17
1280x720	60.44
960x540	64.95
768x432	58.09
640x360	49.38

Ratio	MAC
x2	514M
x3/2	520M
x1	1150M
x3/4	182M
x1/2	122M
x2/5	174M
x1/3	144M

OUR ROADMAP

iSIZE BitSave

1. Mobile & Wireless: **Save bandwidth**
   - Able to stream at lower signal strength (signal to noise ratio - SNR)
   - Improve user experience by decreasing fluctuation in video quality when wireless SNR changes
   - Improve data usage on capped connections
   - Improve device battery life by more efficient encoding/decoding
2. **Reduce CDN cost** for content providers
   - 30% less per media asset
3. **Improved live stream experience** and **longer battery life**

iSIZE UpScale

1. Mobile & Wireless: **Save bandwidth**
2. **Enhancing content** on the client side (HD→4K, 4K→8K)
3. **Enhancing camera capture** (HD camera upgraded to 4K via software)
4. **Enhancing digital zoom** with deep learning

iSIZE BitMind (Under development)

1. **Intelligent cropping of bitstreams** for extreme bandwidth and complexity savings when the "viewer" is a DeepNet
2. **1000x speedup** against conventional DeepNets in video
3. **20x reduction in bitrate** (video streams down to 3kbps!)
4. **Suitable for AI** on wearables, robotics, self-driving vehicles and UAVs

Backwards compatible / advanced machine learning

BEYOND BITSAVE AND THE CURRENT SOLUTIONS

FUTURE POTENTIAL AND PAPERS IN THE PUBLIC DOMAIN

"Deep video precoding" preprint (also presented at IBC 2019), see: https://arxiv.org/pdf/1908.00812.pdf

Use compressed-domain info from the codec for superfast video tagging, action recognition, or fast event detection, e.g., see: https://arxiv.org/abs/1710.05112

Make further rate saving by determining the bitstream subsets needed by machine learning systems, e.g., stream only what a deep neural network needs from the video in order to carry out reliable analysis/classification, see our recent preprint: https://arxiv.org/abs/1810.03964

Use the speed and bandwidth efficiency of these solutions, to allow for real-time analytics on video for wearable cameras, robotic applications, etc.

Build compact video signatures based on codec bitstream data for retrieval of similar videos for recommendation services and linked data analytics

“QUALITY MEANS DOING IT RIGHT
WHEN NO ONE IS LOOKING”

Henry Ford