iSIZE BitSave mk3-lite for real-time communications

Summary: In the iSIZE BitSave mk3-lite model, we significantly reduce the bitrate required to encode high-quality conversational video (1080p at 25/30fps) without any sacrifice in visual quality of the streamed videos. This is achieved with a single-frame, encoder-agnostic, preprocessing stage prior to encoding with any built-in encoder (AVC, HEVC, VP9, AV1, VVC, etc.). No change in encoding, streaming or decoding side is required. The overall architecture is shown in Fig. 1. Results with VMAF (as a standard objective quality metric) as well as ITU-T P.910 DCR tests show that at the top-range of quality (MOS>4.5), 43%-53% saving is achieved over HEVC and VVC while the objective and subjective quality remains on-par to that of the encoder or even increases slightly. In terms of hardware utilization for real-time processing for 1080p@30fps:

- Only 10% of a commodity NVIDIA GTX GPU is used;
- on mainstream Intel i9/i5 CPUs, less than 20% of the CPU is used (51% utilization for Intel i5, which drops to 40% for 720p@30fps);
- on a Qualcomm Snapdragon 888 GPU, 38% of the GPU is used (18% for 720p@30fps).

Such hardware is very widely available in the retail market today.

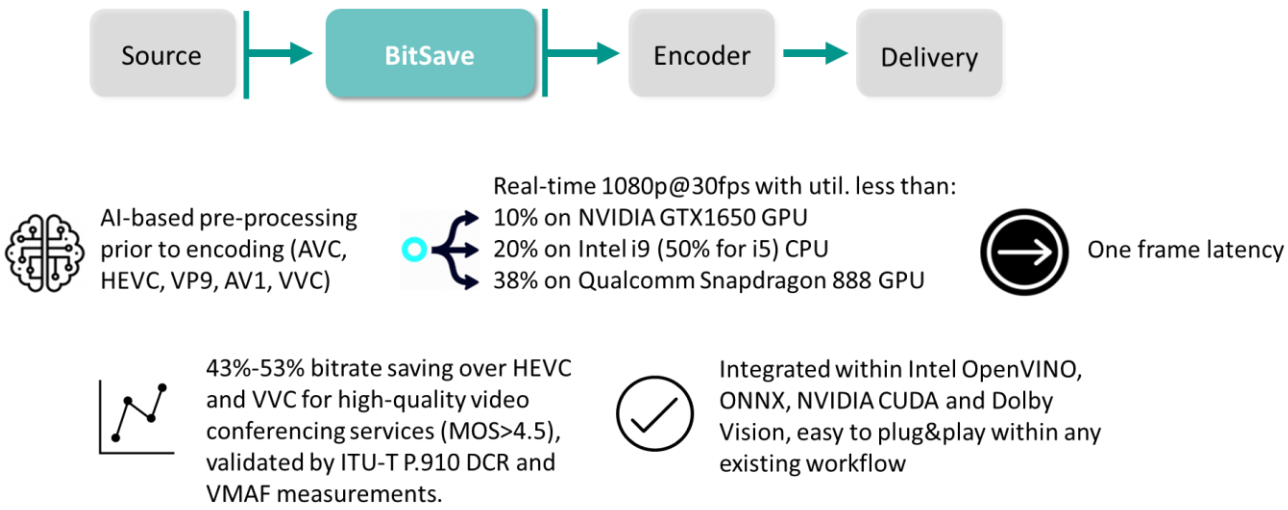

Fig. 1. Processing flow and key aspects of BitSave. The content is preprocessed with a single pass per frame prior to encoding. No change in encoding or decoding side is required.

P.910 DCR tests: In order to be able to assess the perceived quality improvement from the use of iSIZE BitSave mk3-lite, we have carried out ITU-T P.910 DCR tests with a 5-scale rating in collaboration with BBC R&D and Queen Mary University of London. The test utilized 24 raters that have been admitted after eyesight and color-blindness tests. The ratings have been post-processed with the state-of-the-art SUREAL package from Netflix to ensure rater ambiguity and bias is taken into account, see the Netflix SUREAL repo for more information: https://github.com/Netflix/sureal

Sixteen diverse video conferencing clips were used. Sample screenshots of some of the clips can be seen in Fig. 2. All clips were obtained from Creative Commons sources (pexels.com) at high-quality 1080p resolution. The content was assessed in two versions:

- Sources are encoded with crf={23,26,37,46} for HEVC x265 preset=veryslow and crf={23,27,43,53} with VVC vvenc v1.0, preset=medium.

- Sources are preprocessed by the iSIZE BitSave mk3-lite model and then encoded with crf={26,29,39,47} for HEVC x265 preset=veryslow and crf={27,30,46,55} with VVC vvenc v1.0, preset=medium.

The crf encodings are chosen based on computing VMAF scores between sources and preprocessed clips after encoding, so that test sequences with similar VMAF scores are used. This aligns the quality expectation for BitSave+{HEVC,VVC} to {HEVC,VVC} for each of the three quality-bitrate points under consideration. Indeed, the P.910 results show that the average recovered quality scores (MOS) from the raters turn out to be equivalent for BitSave+{HEVC,VVC} vs. {HEVC,VVC}.

Fig. 2. Example screenshots from some of the utilized clips for ITU-T P.910 DCR rating of the {HEVC,VVC} encodings with and without BitSave. The duration of each test clip was ~10 seconds.

The set of average recovered quality scores (RQS) after SUREAL processing is shown in Fig. 3 and Fig. 4 for HEVC and VVC, respectively. The figures show the results for {HEVC,VVC} and BitSave+{HEVC,VVC} for the four CRFs used. In addition, Fig. 5 and Fig. 6 show the corresponding VMAF results, which are measured using the original input videos as reference. By comparing Figs. 3-4 and Figs. 5-6, we observe there is a good alignment between the P.910 subjective RQS results and the objective VMAF scores. This indicates that the reported savings of 43%-53% for the top-range of quality (RQS>4.5) are validated both objectively and subjectively by the controlled conditions of the P.910 DCR test. Tables 1 & 2 show the breakdown of the achieved savings per clip for the top-quality region (RQS>4.5), which is the region of commercial relevance to high-quality video conferencing services. A comparison of the test clips for the top-quality VVC encoding is found here:
https://www.dropbox.com/sh/qwt2ejg8cn9nm10/AAB5Uoy3GSGTKbyXRWtoolRya?dl=0

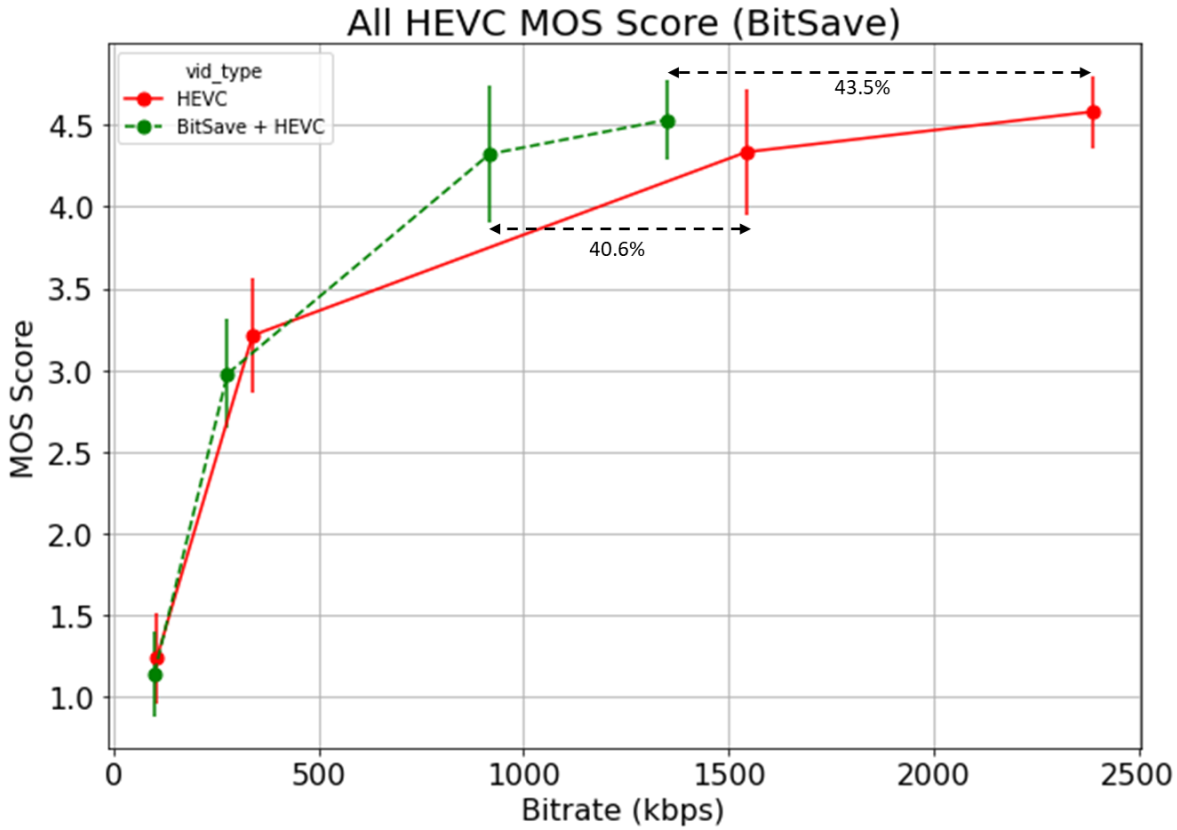

Fig. 3. Average of Recovered Quality Scores (MOS) for all test clips after SUREAL processing of HEVC and BitSave+HEVC.

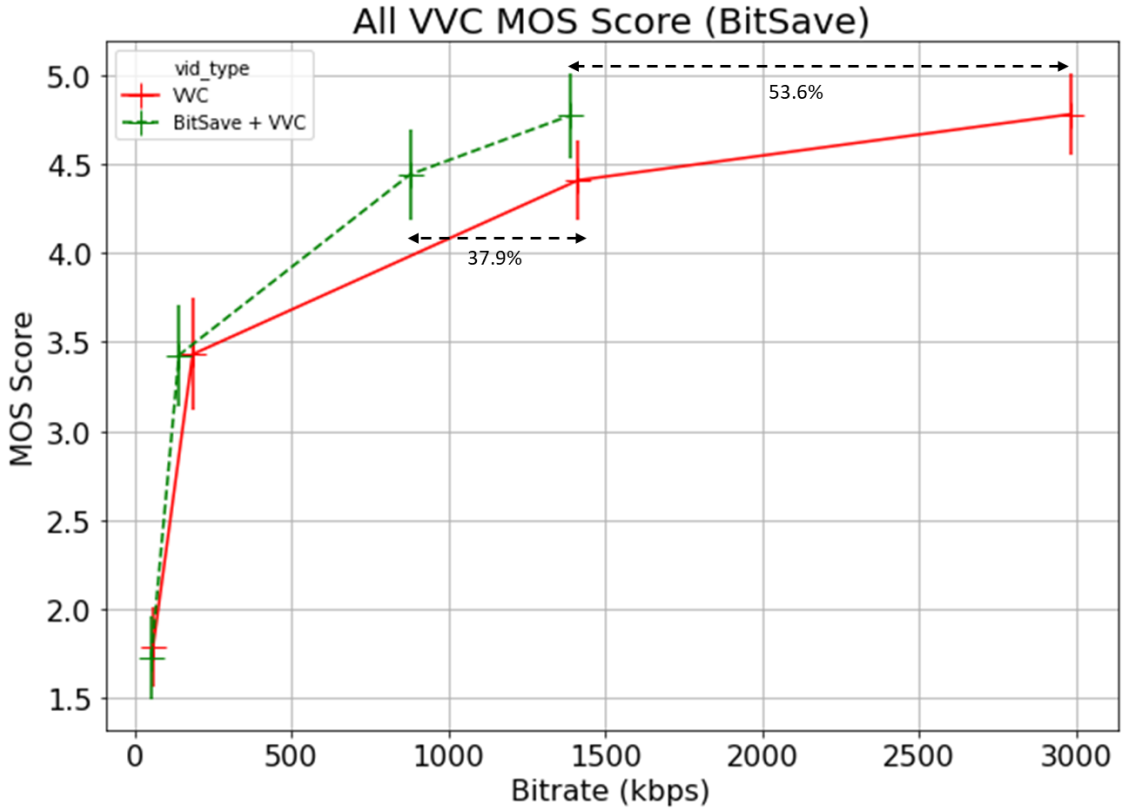

Fig. 4. Average of Recovered Quality Scores (MOS) for all test clips after SUREAL processing of VVC and BitSave+VVC.

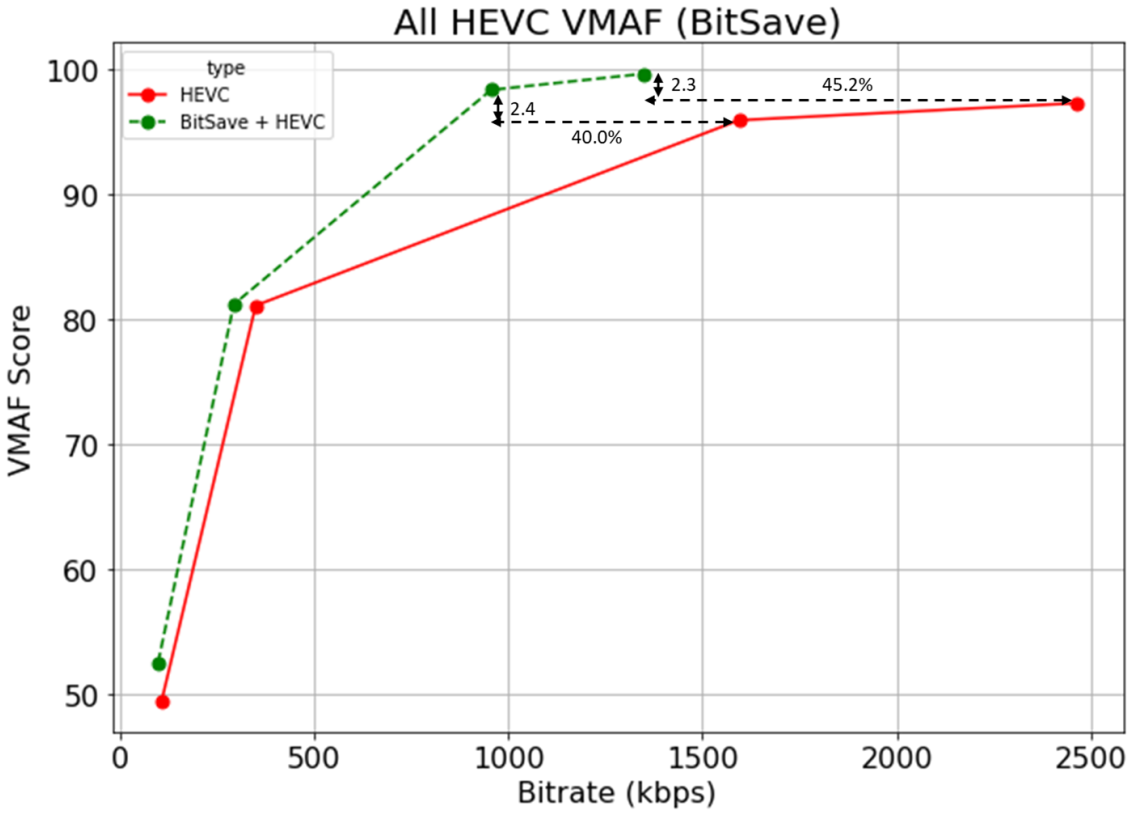

Fig. 5. Average VMAF of HEVC and BitSave+HEVC.

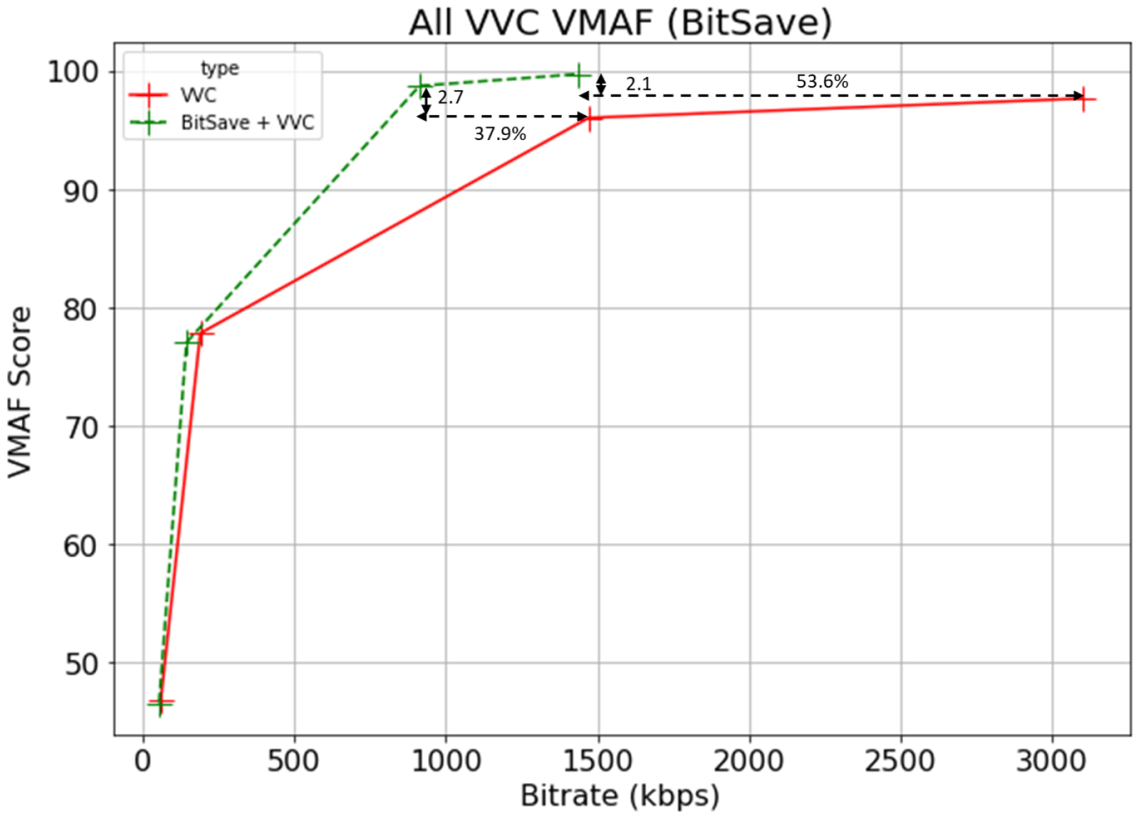

Fig. 6. Average VMAF of VVC and BitSave+VVC.

Table 1. Percentile saving per sequence for the top CRF of VVC (vvenc), which corresponds to the high-quality regime (RQS>4.5).

Video Sequence Name	Percentile Saving (%)	Recovered Quality Score (MOS)	
		VVC	BitSave + VVC
man_talking_in_front_of_office_window	56	4.9	4.9
pexels_roberto_hund_5354705	54	4.2	4.7
portrait-woman-walking-nice-backstreet	57	4.7	4.9
psychologist-taking-notes-with-man-talking_4k	52	4.8	4.8
santa-claus-having-a-video-chat	50	5.0	4.8
sign-language-at-coffee-table_4k	49	4.8	4.5
soldiers-in-therapy-meeting_1080p	52	4.9	4.6
three-people-outdoor-with-glare-4K	54	4.1	4.8
tv-reporter-at-crime-scene	54	4.9	4.8
two-men-having-a-sandwich-break_4k	51	4.9	4.3
two-scientists-with-massive-touchscreen_4k	51	4.8	4.8
woman_sit_in_chair_with_laptop	55	4.8	4.9
woman_talk_in_city_slowmo	57	5.0	5.0
woman_talking_with_earplugs	53	4.9	4.9
woman-at-table-handing-phone-to-man_4k	59	4.8	4.7
woman-in-front-of-grass-wall-talking-to-man	53	4.9	4.8
Average	**53**	**4.8**	**4.8**

Table 2. Percentile saving per sequence for the top CRF of VVC (vvenc), which corresponds to the high-quality regime (RQS>4.5).

Video Sequence Name	Percentile Saving (%)	Recovered Quality Score (MOS)	
		HEVC	BitSave + HEVC
man_talking_in_front_of_office_window	48	4.6	4.5
pexels_roberto_hund_5354705	46	4.6	4.5
portrait-woman-walking-nice-backstreet	45	4.2	4.4
psychologist-taking-notes-with-man-talking_4k	39	4.5	4.5
santa-claus-having-a-video-chat	38	4.5	4.8
sign-language-at-coffee-table_4k	38	4.5	4.3
soldiers-in-therapy-meeting_1080p	37	4.6	4.5
three-people-outdoor-with-glare-4K	42	4.4	4.7
tv-reporter-at-crime-scene	38	4.5	4.7
two-scientists-with-massive-touchscreen_4k	32	4.6	4.4
woman-at-table-handing-phone-to-man_4k	37	4.8	4.5
woman-in-front-of-grass-wall-talking-to-man	44	4.8	4.6
Average	**40**	**4.6**	**4.5**

Runtime cost: We qualified the runtime performance of iSIZE BitSave mk3_lite on:

- A commodity NVIDIA GTX1650 GPU;
- Intel i9 and i5 CPUs;
- a Qualcomm Snapdragon 888 GPU.

These cover major product lines of retail laptops, e.g., Dell XPS, Alienware, MSI, Acer, ASUS, HP, Razer, Origin, Lenovo, Apple (pre-2019), etc. All of these commodity devices, as well as all modern smartphones with a Qualcomm Snapdragon GPU, will easily be able to support BitSave.

Typical runtime performance numbers for NVIDIA are found in Table 2. The benchmarks show that less than 10% utilization is required for 1080p, and 1% or less for 720p or lower-resolution video. Memory requirements are also very modest, needing less than 0.5GB for BitSave mk3_lite.

Table 3. GPU memory, utilization and runtime of MK3-lite model on NVIDIA GeForce GTX (average of 1000 runs per resolution on a Dell XPS15 laptop).

Measurements on NVIDIA GeForce GTX1650 Ti from Live Webcam Feed				
Resolution	Memory Usage (GB)	Utilization (%)	Time per Frame (ms)	FPS
1080p	0.5	9	34	29
720p	0.1	1	35	29
480p	0.1	~1	32	31

Table 4. CPU memory, utilization and runtime of MK3-lite model (average of 1000 runs per resolution) on: (a) Intel i9 on a Dell XPS 15 laptop; (b) Intel i5 CPU on an HP Pavilion 13.

Measurements on CPU [Intel i9-10885H@2.40GHz] from Live Webcam Feed				
Resolution	Memory Usage (GB)	Utilization (%)	Time per Frame (ms)	FPS
1080p	0.6	19	34	30
720p	0.3	9	34	30
480p	0.1	4	34	30

(a)

Measurements on CPU [Intel i5-1035G1@1.00GHz] from Live Webcam Feed				
Resolution	Memory usage (GB)	Utilization (%)	Time per Frame (ms)	FPS
1080p	0.5	51	41	24
720p	0.2	40	33	30
480p	0.1	18	34	30

(b)

(note that 1080p is at 24FPS as this was the webcam FPS on the test machine)

Finally, the corresponding results for Qualcomm Snapdragon 888 are shown in Table 4. Similarly, BitSave can be supported with modest memory and GPU utilization for all resolutions up to 1080p@30fps.

Table 5. GPU memory, utilization and runtime of MK3-lite model on Qualcomm Snapdragon.

Measurements obtained on Snapdragon 888 from Live Webcam Feed (Samsung S21)				
Resolution	Memory usage (GB)	Utilization (%)	Time per Frame (ms)	FPS
1080p	0.4	38	33	30
720p	0.2	18	33	30
480p	0.1	10	33	30