Encoder benchmarking checklist

Same VMAF, different bitrate / speed — one-page operator's guide

Companion to article 6.5 of Block 6. Print on A4. Hand to procurement or run yourself. Fora Soft Learn · v1.0 · 2026-05-17

1. The eight-step recipe

#	Step	What to lock down	Output
1	**Corpus**	6-12 clips: talking head, sports, cartoon, dark, nature	YUV 4:2:0 sources
2	**Encoders**	2-6 encoders, 1-2 presets each, locked	Preset list
3	**Rate points**	4-6 CRFs or fixed bitrates per encoder x clip	CRF/bitrate grid
4	**Encode**	Log wall-clock time + output bitrate	Encode logs
5	**Measure VMAF**	libvmaf, official model (vmaf_v0.6.1 or phone/4K)	Per-frame VMAF
6	**Plot RD**	Log-x bitrate, linear-y VMAF, one curve per encoder	RD curves
7	**BD-rate**	Bjontegaard cubic spline, per encoder pair, per clip	BD-rate %
8	**Aggregate**	Report mean AND per-clip range, not just the mean	Final number

2. Eight ways benchmarks lie — rule each one out

- [ ] One clip only -> always 6+. Per-clip BD-rate range tells you more than the mean.
- [ ] One preset -> match preset speeds across encoders before comparing quality.
- [ ] No quality control -> bitrate without VMAF is meaningless.
- [ ] Arithmetic-mean VMAF only -> also report harmonic mean + 1st percentile.
- [ ] Default encoder settings -> use published reference configs (tune=0, lookahead=120, ...).
- [ ] Wrong VMAF model -> 1080p TV (default) vs phone vs 4K shift VMAF by 3-5 points.
- [ ] Rate-control mode mismatch -> CRF vs CBR vs capped-CRF are not like-for-like.
- [ ] Cherry-picked aggregation -> median hides the worst case, mean hides variance.

3. The reporting table you must produce

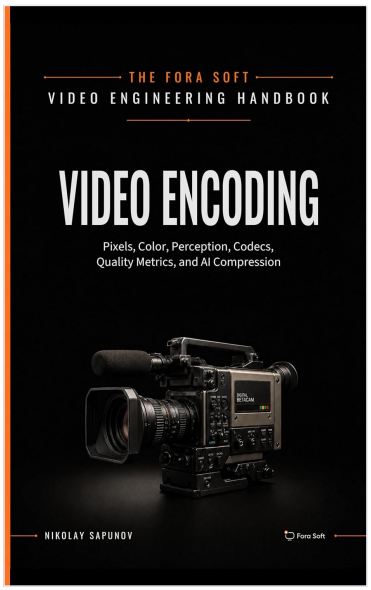
Encoder	Preset	BD-rate vs ref	1080p fps	Watts/stream	Verdict
x264	medium	0% (ref)	120	~	Universal
x265	medium	-30..-35%	30	~	OTT today
SVT-AV1	preset 8	-40..-45%	25	~	Live AV1
SVT-AV1	preset 4	-55..-60%	10	~	Premium VOD
vvenc	medium	-55..-65%	2	~	Future-facing

4. VMAF reporting thresholds (apply both, not just one)

- Harmonic mean VMAF >= 95 -> top-rung perceptual quality.
- 1st percentile VMAF >= 89 -> 99% of frames clear the visible-artefact line.
- Standard deviation reported alongside, not hidden.
- Two encoders at same mean VMAF but different stddev look very different on TV.

5. The five questions for any encoder vendor

1. What corpus did you benchmark on? Was it AOMedia CTC, UVG, MSU, or 'internal'?
2. What is your BD-rate vs x264 medium and x265 medium - on MY content, not yours?
3. What is your 1080p encoded fps on a 16-core CPU at the preset you recommend?
4. Which VMAF model? Mean, harmonic mean, or percentile?
5. Will you run this benchmark live on my reference clip during the procurement call?



GO DEEPER · VOLUME 1 OF 7

Video Encoding: A Complete Guide to Video Compression

From Pixels and Color to H.264, HEVC, AV1, Quality Metrics, and AI

By Nikolay Sapunov, CEO at Fora Soft · 1,522 pages · Kindle & paperback on Amazon

Read it on Amazon → amazon.com/dp/B0HF52B65R

This checklist gives you the eight-step recipe for an honest encoder benchmark and the questions for vendors. Volume 1 covers the science underneath: rate-distortion curves, BD-rate, VMAF models and pooling, encoder presets and rate control, and why one clip or one preset proves nothing.

WHAT YOU WILL FIND INSIDE

- Rate-distortion curves and BD-rate
- Presets and rate control across encoders
- VMAF models, pooling and percentiles
- Building a representative test corpus

THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI
Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF
Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery
Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing
Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs
Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video
Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

Evaluating encoders or a vendor's compression claims?

We will run a same-VMAF benchmark on your own content and give you BD-rate, speed and cost per stream for each candidate. 250+ video products shipped since 2005.

Book a 30-min call → **WhatsApp →** **Email us →**