

Companion to article 5.3 of Block 5. Print on A4.

1. Four-question decision flow (walk top to bottom)

- Q1. One-off VOD master at maximum quality? YES -> software (x265 / SVT-AV1 / x264)
 Q2. Sub-second end-to-end latency required? YES -> GPU + CPU media engine (NVENC + QSV / VideoToolbox)
 Q3. More than 50 concurrent live channels? YES -> VPU / ASIC (NETINT Quadra T1U / T2A or AMD Alveo MA35D)
 Q4. AI inference on the same chip? YES -> GPU (NVENC / VCN) ; otherwise -> GPU, cheapest tier

2. Concurrent real-time 1080p30 HEVC streams per device (2026 typical numbers)

Device	Streams	Power	Strength
CPU dual-socket 64-core (libx265 medium)	~10	~600 W	Best quality per bit
Consumer GPU NVENC (RTX 5070-class)	~12	~220 W	Cheap and ubiquitous
Data-center GPU NVENC (NVIDIA L40S)	~30	~300 W	Multi-NVENC and AI co-located
Intel Arc B580 (QSV / Battlemage)	~12	~190 W	Cheapest AV1 hardware
AMD RX 9070 XT (VCN 5.0)	~14	~300 W	B-frame AV1, competitive quality
Apple M4 Max (VideoToolbox)	~6	~70 W	Low-power HEVC + ProRes
NETINT Quadra T1U VPU	~16	~40 W	Highest streams per watt
NETINT Quadra T2A VPU (dual G5)	~32	~75 W	Datacenter scale, AV1 native

3. The four command lines you actually need

NVENC (NVIDIA) : `ffmpeg -hwaccel cuda -hwaccel_output_format cuda -i in.mp4 \`
 `-c:v h264_nvenc -preset p5 -tune hq -cq 23 out.mp4`
 QSV (Intel) : `ffmpeg -hwaccel qsv -hwaccel_output_format qsv -i in.mp4 \`
 `-c:v hevc_qsv -preset medium -global_quality 23 out.mp4`
 VideoToolbox : `ffmpeg -i in.mov -c:v hevc_videotoolbox -q:v 50 -tag:v hvc1 out.mp4`
 AMF (AMD) : `ffmpeg -i in.mp4 -c:v av1_amf -quality quality -qp_i 24 -qp_p 26 out.mp4`

4. Quality vs speed - rules of thumb (1080p, same bitrate)

x265 medium = baseline VMAF (top performer at slow presets)
 SVT-AV1 preset 6 = +1 to +2 VMAF over x265 medium, ~2x slower than x264
 NVENC HEVC p5 = -1 to -2 VMAF vs x265, 20x to 50x faster
 QSV HEVC medium = -1 to -3 VMAF vs x265, similar speed to NVENC
 VCN 5.0 AV1 = -1 to -2 VMAF vs SVT-AV1 preset 8, hardware-fast
 VideoToolbox HEVC = -2 to -4 VMAF vs x265, lowest power on Apple Silicon
 NETINT Quadra AV1 = -1 to -3 VMAF vs SVT-AV1, real-time at 16+ streams per card

5. Common pitfalls (and the fix)

Pitfall	Fix
Distro FFmpeg has no NVENC	Install BtbN static build or compile with <code>--enable-nvenc</code>
Comparing encoders on VMAF only	Pair VMAF with PSNR and one subjective pass
AV1 hardware live stream stalls	Pin H.264 / HEVC for live, re-encode AV1 in a 2nd pass
Apple HEVC silent in Safari	Add <code>-tag:v hvc1</code> (default hev1 is rejected)
Saturated QSV on conferencing host	Profile end-to-end latency, not just encode speed
Buying VPU below 50 streams	GPU is cheaper at that scale - VPU ROI starts at 50+
Pre-2023 NVENC session limit	Update driver to 522.25 or newer; cap is gone

6. Pick by workload (one line each)

Live conferencing -> QSV / VideoToolbox on client, NVENC on SFU (latency first).
 Live broadcast / OTT -> NETINT Quadra T1U or T2A above 50 channels; NVENC L40S below.
 Per-title VOD library -> SVT-AV1 software for catalogue; NVENC AV1 UHQ for fast tier.
 One-off cinema master -> x265 slow or SVT-AV1 preset 4 on CPU; never hardware.
 AI + encoding fused -> NVIDIA GPU; NETINT's 36 TOPS NPU for ROI / content-adaptive RC.