

Companion to article 4.1 of Block 4. Print on A4 or US Letter.

The nine stages, in order

- | | |
|-------------------------------------|--|
| 1. Partitioning | Split frame into blocks. Larger for flat regions, smaller for detail. |
| 2. Prediction | Intra (from same frame) or Inter (from past frame + motion vector). |
| 3. Residual | Original minus prediction. Most bits in a stream describe residuals. |
| 4. Transform | DCT / ADST. Rotates residual into frequency coefficients. Lossless. |
| 5. Quantization | Divide each coefficient by Qstep. *** Only stage where info is lost. *** |
| 6. Entropy coding | CABAC / multi-symbol arithmetic. Lossless final pack. |
| 7. Inverse quant + transform | Encoder runs decoder inside itself to stay aligned. |
| 8. In-loop filter | Deblocking + SAO/CDEF/ALF. Smooths block edges on the reference. |
| 9. Reference buffer | Decoded Picture Buffer - feeds future inter-predictions. |

What changed in each generation

Stage	MPEG-2 (1994)	H.264 (2003)	HEVC (2013)	AV1 (2018)
Block size	16x16 fixed	16x16 + 8x8 + 4x4	CTU 64x64 quad-tree	Superblock 128x128
Intra modes	DC only	9 modes	35 angular	95 (incl. CfL, Paeth)
Motion precision	Half-pel	Quarter-pel	Quarter-pel	Eighth-pel
Reference frames	1-2	Up to 16	Up to 16	Up to 8 + compound
Transform	8x8 DCT	4x4 / 8x8 int. DCT	4x4 to 32x32 DCT+DST	4x4 to 64x64 + ADST
Entropy coder	VLC	CAVLC or CABAC	CABAC	Multi-symbol arith.
In-loop filters	None	Deblocking	Deblock + SAO	Deblock + CDEF + LR
1080p30 typical	8-15 Mbps	3-5 Mbps	1.5-3 Mbps	0.8-1.6 Mbps

Common gotchas when reasoning about codecs

- Standard != encoder** H.264 specifies bitstream syntax, not encoder behaviour. Two H.264 encoders produce different files.
- Quant is the only lossy step** Everything else - prediction, transform, entropy - is invertible. Tune Qstep, not the rest.
- Encoder runs a decoder inside** Reconstruction loop keeps encoder and decoder pixel-identical across thousands of frames.
- In-loop filters compound** They improve the reference, not the displayed frame. Gains accumulate across the GOP.
- RDO dominates encoder CPU** $J = D + \lambda R$. Search space grows fast with each generation; that is where minutes go.
- Decoder is always cheaper** Standards push decoder complexity down so phones can play 4K at 1x real time.