moz://a

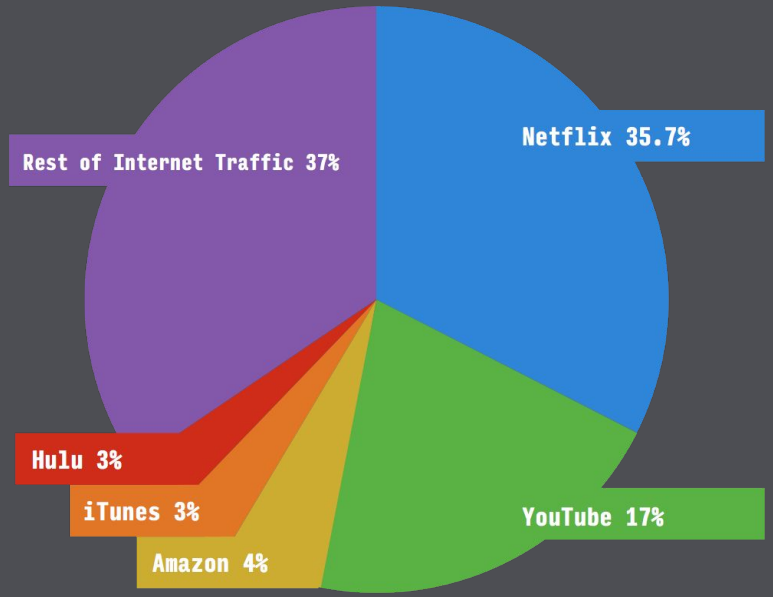
Into the Depths: The Technical Details Behind AV1

Nathan Egge <negge@mozilla.com>

Mile High Video Workshop 2018

July 31, 2018

North America Internet Traffic



82% of Internet traffic by 2021 Cisco Study

Alliance for Open Media (AOM)

Goals of the Alliance:

- Produce a video codec for a broad set of industry use cases
  - Video on Demand / Streaming
  - Video Conferencing
  - Screen sharing
  - Video game streaming
  - Broadcast
- Open Source and Royalty Free
- Widely supported and adopted
- At least 30% better than current generation video codecs

AV1 Coding Tools Overview

- New high-level syntax
  - Easily parsed sequence header, frame header, tile header, etc
- New adaptive multi-symbol entropy coding
  - Up to 16 possible values per symbol
- New coefficient coder
  - LV-MAP exploits multi-symbol arithmetic coder
- More block sizes
  - Prediction blocks from 128x128 down to 4x4
    - Rectangular blocks
      - 1:2 and 2:1 ratios (4x8, 8x4, etc)
      - 1:4 and 4:1 ratios (4x16, 16x4, etc)
  - Transform sizes from 64x64 down to 4x4
    - Includes rectangular transforms 1:2, 2:1 and 1:4, 4:1 ratios
- More transform types
  - 16 possible transform types
    - Row and column chosen from: IDTX, DCT, DST, ADST
- More references
  - Up to 7 per frame (out of a store of 8)
- Spatial and temporal scalability
- Lossless mode
- Chroma subsampling
  - 4:4:4, 4:2:2, 4:2:0, monochrome
- More prediction modes
  - Intra
    - 8 main directions plus delta for up to 56 directions
    - Smooth HV modes interpolate across block
    - Palette mode with index map up to 8 colors
    - Chroma from Luma intra predictor
    - Intra Block Copy
  - Inter
    - Expanded reference list (up to 7 per frame)
    - Allow ZEROMV predictor, which isn't always (0,0)
    - Compound mode
      - Inter-Intra prediction
        - Depends on difference between pixel prediction
        - Smooth blending limited to certain intra modes
      - Wedge codebook (Inter-Inter, or Inter-Intra)
    - Warped motion local affine model with neighbors
    - Global motion affine model across entire frame
- Loop filtering
  - Deblocking filter
  - Constrained Directional Enhancement Filter
  - Loop restoration
- Film grain synthesis

Full AV1 Specification: https://aomediacodec.github.io/av1-spec/

AV1 Coding Tools Overview

- New high-level syntax
  - Easily parsed sequence header, frame header, tile header, etc
- New adaptive multi-symbol entropy coding
  - Up to 16 possible values per symbol
- New coefficient coding
  - LV-MAP exploits multi-symbol arithmetic coder
- More block sizes
  - Prediction blocks from 128x128 down to 4x4
    - Rectangular blocks
      - 1:2 and 2:1 ratios (4x8 & 8x4, etc)
      - 1:4 and 4:1 ratios (4x16, 16x4, etc)
  - Transform sizes from 64x64 down to 4x4
    - Includes rectangular transforms 1:2, 2:1, and 1:4, 4:1 ratios
- More transform types
  - 16 possible transform types
    - Row and column chosen from: IDTX, DCT, DST, ADST
- More references
  - Up to 7 per frame (out of a store of 8)
- Spatial and temporal scalability
- Lossless mode
- Chroma subsampling
  - 4:4:4, 4:2:2, 4:2:0, monochrome
- More prediction modes
  - Intra
    - 8 main directions plus delta for up to 56 directions
    - Smooth DC/V modes interpolate across block
    - Palette mode with index map up to 8 colors
    - Chroma from Luma intra predictor
    - Intra Block Copy
  - Inter
    - Expanded reference list (up to 7 per frame)
    - Also ZEROMV predictor, which isn't always (0,0)
    - Compound mode
      - Inter-Intra prediction
        - Depends on difference between pixel prediction
        - Smooth blending limited to certain intra modes
      - Wedge codebook (Inter-Inter, or Inter-Intra)
    - Warped motion local affine model with neighbors
    - Global motion affine model across entire frame
- Loop filtering
  - Deblocking filter
  - Constrained Directional Enhancement Filter
  - Loop restoration
- Film grain synthesis

Full AV1 Specification: https://aomediacodec.github.io/av1-spec/

AV1 Coding Tools Overview

- New high-level syntax
  - Easily parsed sequence header, fram[illegible]der, t[illegible] heade[illegible] etc
- New adaptive multi-symbol entropy [illegible]od
  - Up to 16 possible valu[illegible]s per syn[illegible]
- New coefficient co[illegible]
  - LV-MAP exploits m[illegible]lti-sy[illegible]bol ant[illegible]
- More block sizes
  - Prediction blocks from 128x128 do[illegible]n to 4x[illegible]
    - Rectangular blocks
      - 1:2 and 2:1 rat[illegible]s (4x[illegible] & [illegible]4, etc
      - 1:4 a[illegible]d [illegible]1 rat[illegible]s (4x[illegible], 16x[illegible]
  - Transform sizes f[illegible]m 64x[illegible]down
    - Includes rectan[illegible]ular transfo[illegible]ations
- More transform types
  - 16 possible transform type[illegible]
    - Row and column chos[illegible]n from: I[illegible]CT, DST, ADST
- More references
  - Up to 7 per frame (out of a store of 8)
- Spatial and temporal scalability
- Lossless mode
- Chroma subsampling
  - 4:4:4, 4:2:2, 4:2:0, monochrom[illegible]e
- More prediction [illegible]od[illegible]s
  - Intra
    - [illegible]ctio[illegible]s plus delta for up to 56 directions
    - [illegible]modes interpolate across block
    - [illegible]with index map up to 8 colors
    - [illegible]r Lu[illegible]a intra predictor
    - [illegible]x copy
  - Ex[illegible]anded [illegible]ence list (up to 7 per frame)
    - [illegible] ZEROMV predictor, which isn't always (0,0)
    - [illegible]und mode
    - [illegible]prec[illegible]tio[illegible]
    - [illegible]s of diffe[illegible]ence between pixel prediction
    - [illegible]oth blen[illegible]ing limited to certain intra modes
    - [illegible]odebook (Inter-Inter, or Inter-Intra)
    - Wa[illegible]motion local affine model with neighbors
    - Global motion affine model across entire frame
- Loop filtering
  - Deblocking filter
  - Constrained Directional Enhancement Filter
  - Loop restoration
- Film grain synthesis

THIS LIST IS TOO LONG TO COVER IN 30 MINUTES!!!

LET'S TRY

Full AV1 Specification: https://aomediacodec.github.io/av1-spec/

Profiles

Main
- 8-bit and 10-bit
- 4:0:0 and 4:2:0 chroma subsampling

High
- 8-bit and 10-bit
- 4:0:0, 4:2:0 and 4:4:4 chroma subsampling

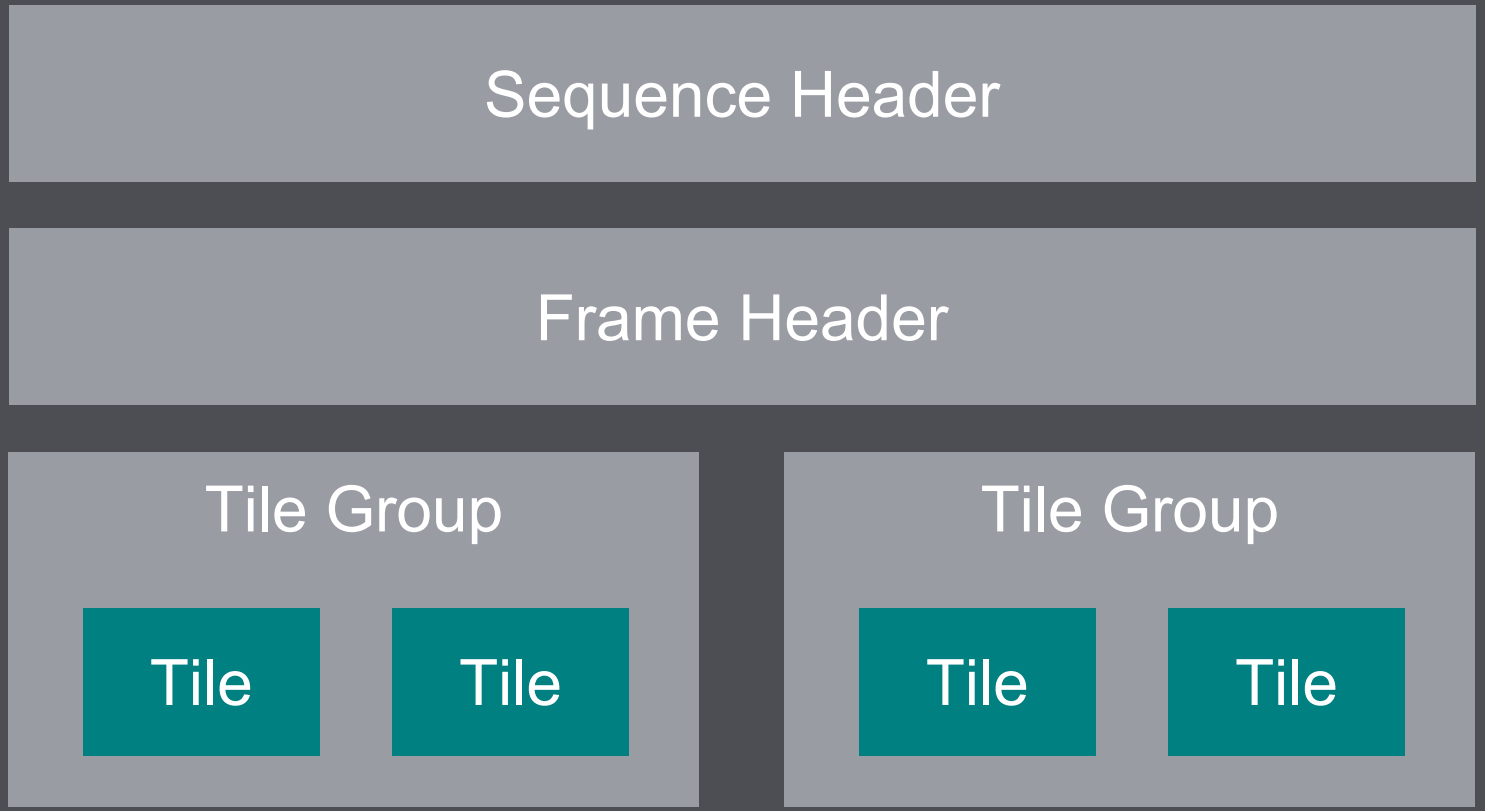
Professional
- 8-bit, 10-bit and 12-bit
- 4:0:0, 4:2:0, 4:2:2 and 4:4:4 chroma subsampling

Levels

For a given sequence, place limits on:

- frame size (width and height)
- maximum picture size (area in samples)
- maximum display rate (samples per second)
- maximum decode rate (samples per second)
- average rate (Mbits per second)
- high rate (Mbits per second)
- maximum number of tiles
- maximum number of tile columns

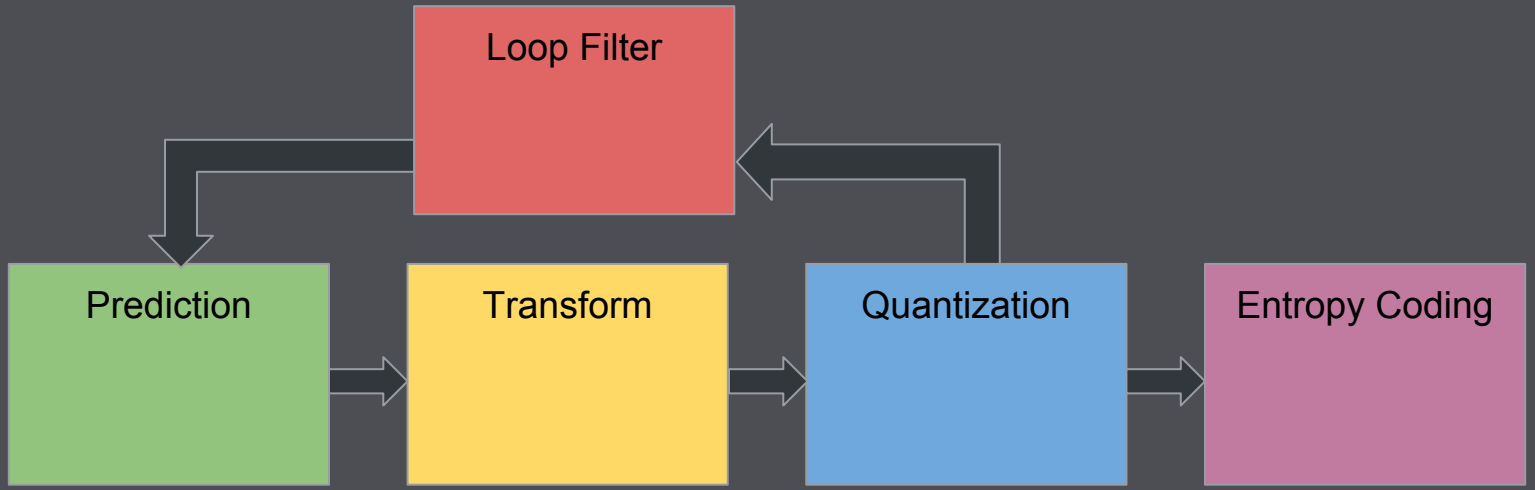
High Level Syntax



Colors and HDR

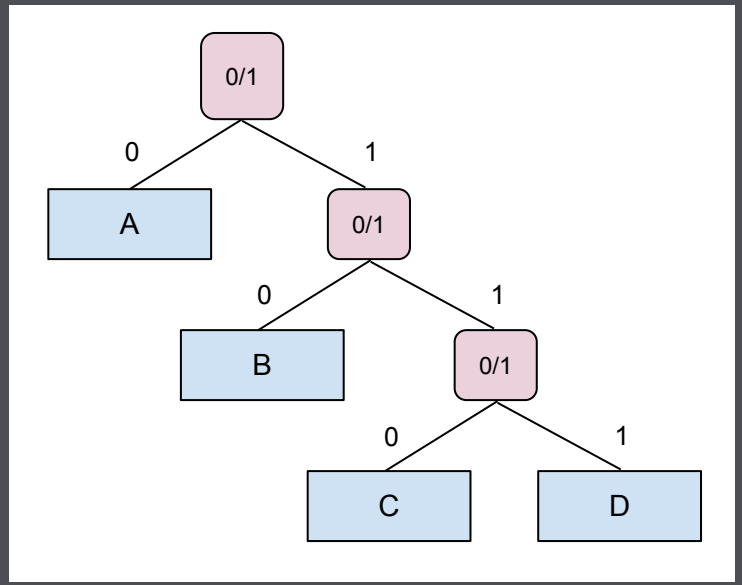
- Colorspace, color matrix, transfer functions, etc. can be encoded directly in the bitstream
  - Chroma siting and levels too
- HDR metadata can be added through the Metadata OBU syntax

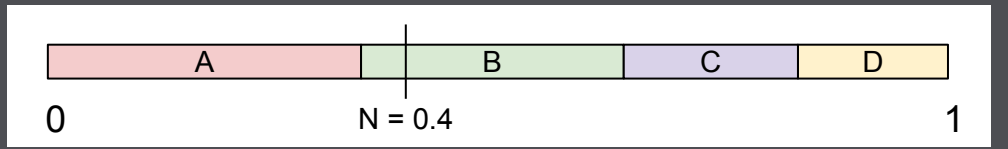
Codecs 101


Loop Filter
Prediction
Transform
Quantization
Entropy Coding

Multi-Symbol Entropy Coder

- Arithmetic Range Coder
- Code both binary symbols and multi-symbols
  - Alphabet sizes up to 16
- Improve EC throughput with high rate streams
  - Instead of 1 bit per cycle, decode up to 4
- Use 8x9 -> 17 bit multiples when coding
  - 15-bit CDFs shifted down before multiply
  - Adaptation still occurs with 15-bit precision
- Fast adaptation mode for first few symbols





Transform Types

VP9 has two types: DCT and ADST

- Chosen independently for horizontal / vertical directions
- Signaled once per prediction block

AV1 has four types:

- DCT
- ADST
- FlipADST (mirror image of ADST)
- Identity (no transform)

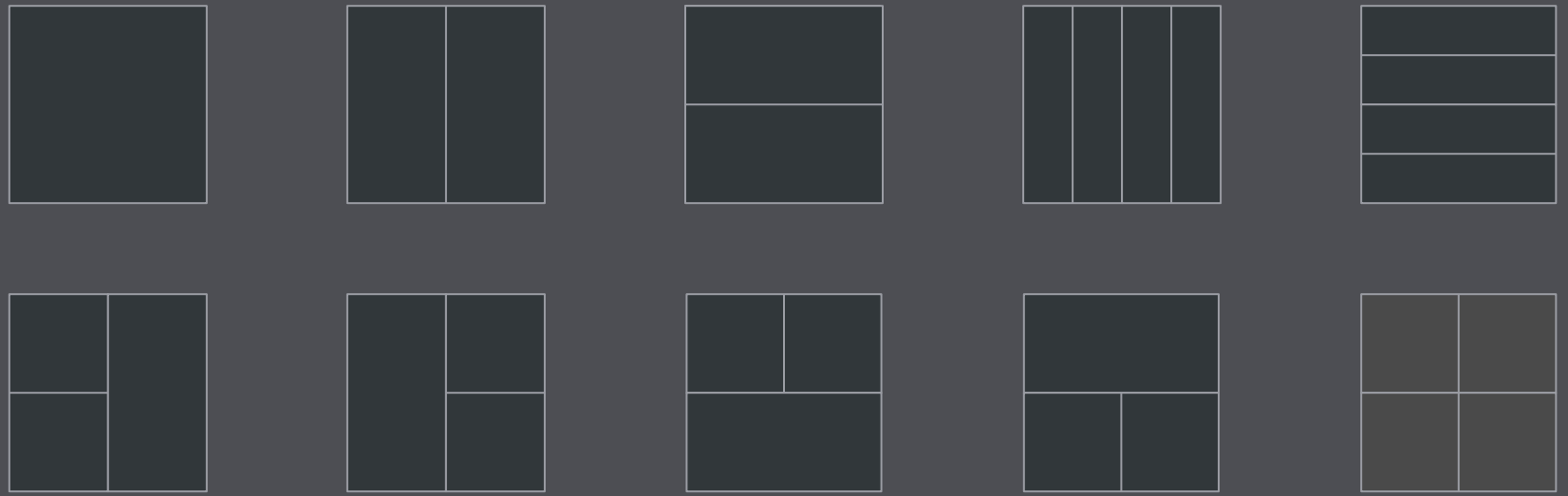
Still chosen independently for horizontal / vertical directions

- Total of 16 possible combinations
- Not all combinations allowed in all contexts (e.g., no FlipADST for intra)

Signaled once per transform block

Prediction Block Structure

10 different splitting modes

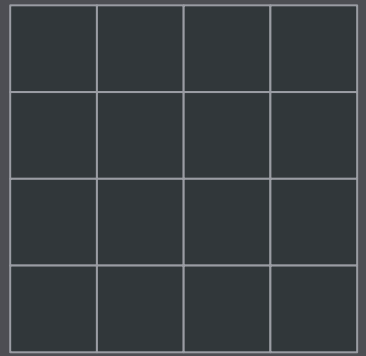


- Last (4-way) split is recursive

Transform Block Sizes: Intra

Signaling mostly unchanged from VP9

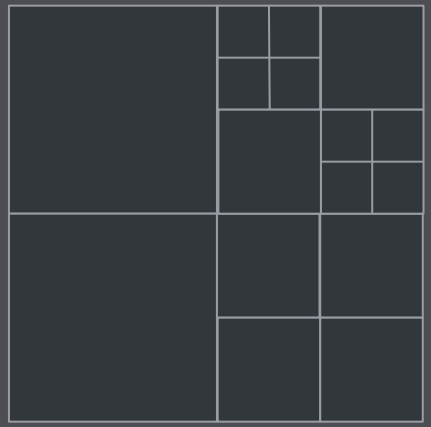
- One transform size per prediction block
- For rectangular prediction blocks, largest rectangular transform that fits allowed, e.g., 1:2, 2:1, 4:1 and 1:4 ratio transform blocks
- Transform sizes go up to 64x64
  - Only upper left 32x32 region allowed to be non-zero



Transform Block Sizes: Inter

Signaling completely different from VP9

- Four way quad tree splitting
- For rectangular prediction blocks, largest rectangular transform that fits also allowed
- Available sizes same as intra

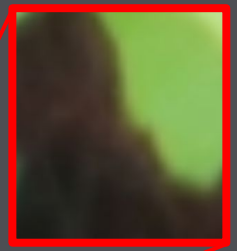


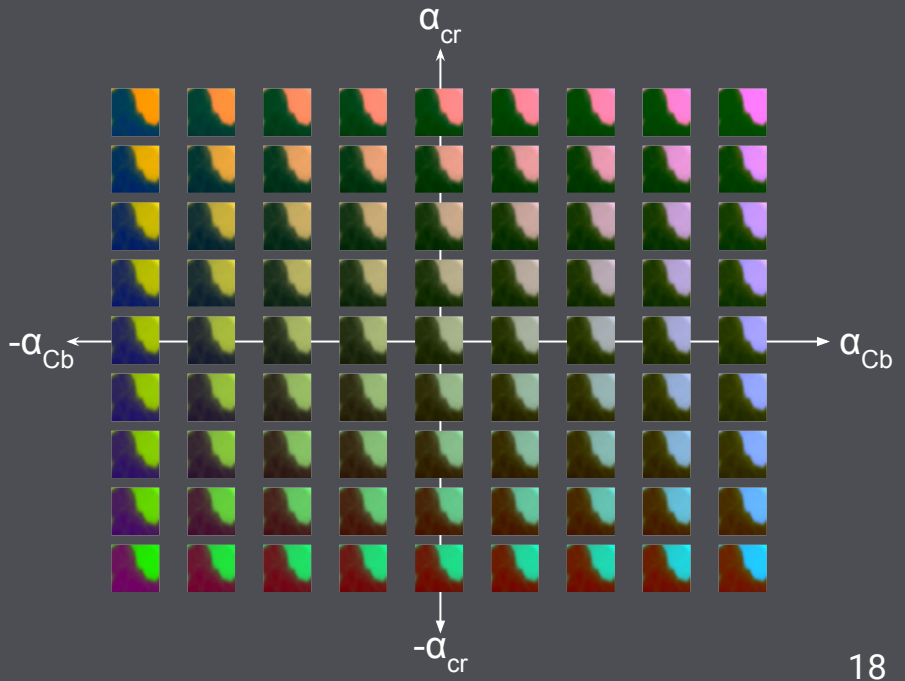
Intra Prediction Modes

- More directional modes
  - 8 main directions plus delta for up to 56 directions
  - Not all modes available at smaller sizes
- Smooth H + V modes
  - Smoothly interpolate between values in left column (resp. above row) and last value in above row (resp. left column)
- Paeth predictor mode
- Palette mode
  - Color index map with up to 8 colors
  - Separate palettes for Y, U and V planes
  - Palette index coded using context model for each pixel in the block
  - Pixels predicted in ‘wavefront’ order to allow parallel computation
- Chroma from Luma

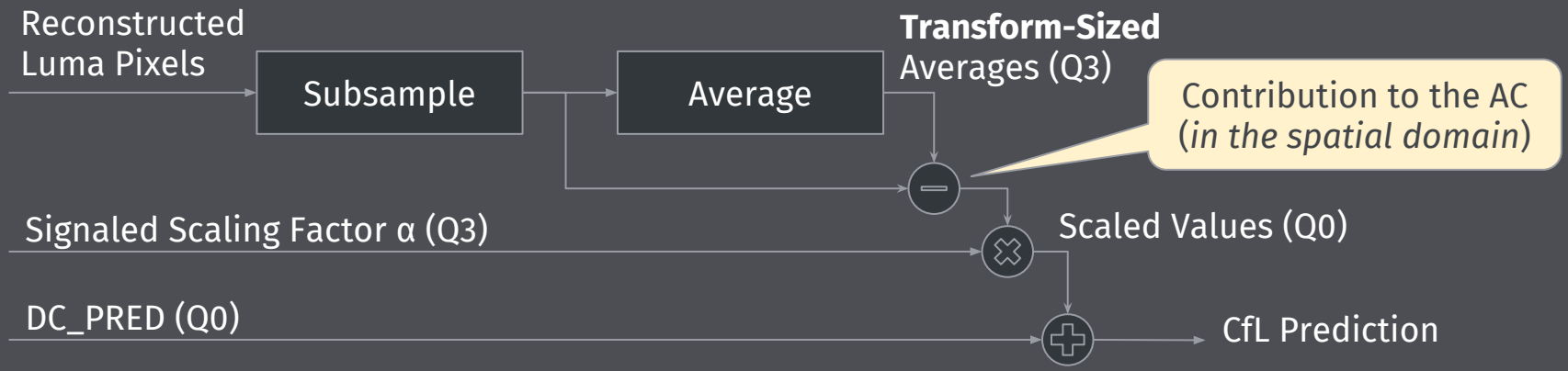
Chroma from Luma Intra Prediction

- Predict chroma channel based on decoded luma
  - Encoder signals best correlation constants: α_{cb} and α_{cr}
- Good for screen content or scenes with fast motion





Chroma from Luma Algorithm



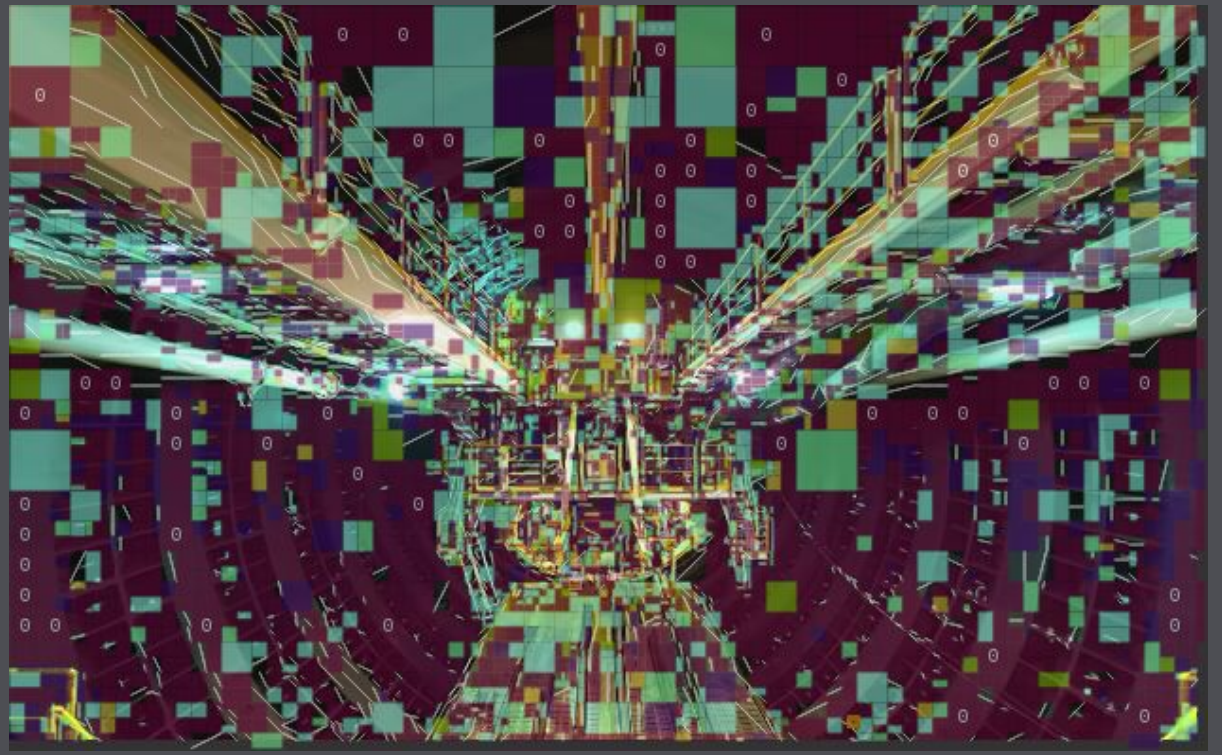
UV Mode Selection Example (https://goo.gl/6tKaB8)

CFL_PRED 17%

DC_PRED 44.36%

TM_PRED 7.98%

SMOOTH_PRED 4.85%



Ohashi0806shield.y4m
QP = 55

Awesome for Gaming (Twitch dataset)

	BD-Rate (%)						
	PSNR	PSNR-HVS	SSIM	**CIEDE2000[1]**	PSNR Cb	PSNR Cr	MS SSIM
Average	-1.01	-0.93	-0.90	**-5.74**	-15.55	-9.88	-0.81

https://arewecompressedyet.com/?job=no-cfl-twitch-cpu2-60frames%402017-09-18T15%3A39%3A17.543Z&job=cfl-inter-twitch-cpu2-60frames%402017-09-18T15%3A40%3A24.181Z

Notable Mentions

	BD-Rate (%)						
	PSNR	PSNR-HVS	SSIM	**CIEDE2000[1]**	PSNR Cb	PSNR Cr	MS SSIM
Minecraft	-3.76	-3.13	-3.68	**-20.69**	-31.44	-25.54	-3.28
GTA V	-1.11	-1.11	-1.01	**-5.88**	-15.39	-5.57	-1.04
Starcraft	-1.41	-1.43	-1.38	**-4.15**	-6.18	-6.21	-1.43

Minecraft

MINECRAFT_10_120f.y4m

GTA V

GTAV_0_120f.y4m

Starcraft

STARCRAFT_10_120f.y4m

Motion Vector Coding

- Each frame has a list of 7 previous frames to reference (out of a pool of 8)
  - Can reference non-displayed frames, so many possible structures
- Construct list of top 4 MVs for a given reference / reference pair from neighboring area
- Complicated entropy coding scheme

Compound Prediction

(½, ½) weights like VP9

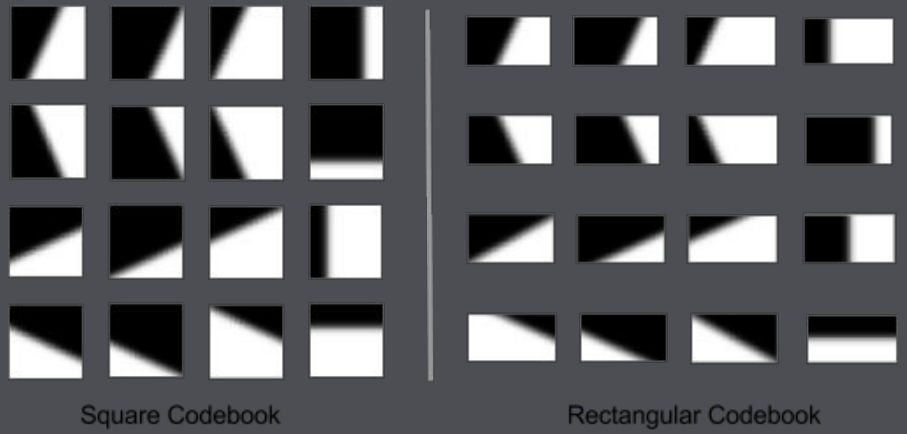
Inter-inter compound segment

- Pixel weights depend on difference between prediction pixels

Inter-intra gradual weighting

- Smoothly blends from inter to intra prediction
- Only a limited set of intra modes allowed (DC, H, V, Smooth)

Wedge codebook (inter-inter or inter-intra)



Global Motion

- Defines up to a 6-parameter affine model for the whole frame (translation, rotation and scaling)
- Blocks can signal to either use the global motion vector or code a motion vector like normal
    - If global motion isn’t used, default is 0,0

Warped Motion

- Use neighboring blocks to define same motion model within a block
  - Decomposed into two shears with limited range
    - Similar complexity to subpel interpolation

Segmentation IDs

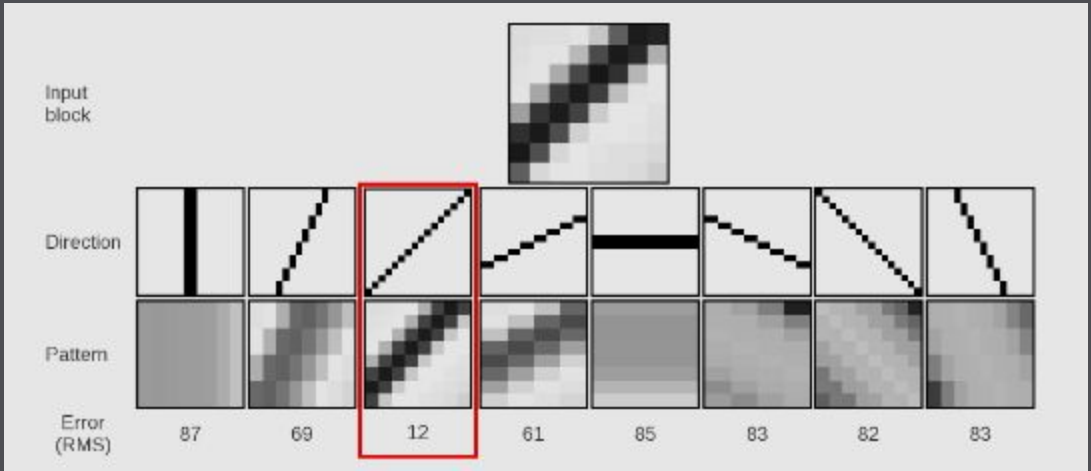
- Up to 8 possible segment labels (3 bits)
  - Value set per label, e.g., filter strength, quantizer, reference frame, skip
  - Signaled per prediction block, down to 8x8
- Can either predict segment ID temporally or spatially (chosen per frame)
  - Spatial prediction
    - Used to change quantizer/loop filter strength
    - Useful for adaptive quantization, e.g., for activity masking
    - Useful for temporal RDO, e.g., MV-tree
  - Temporal prediction
    - Useful for predicting temporal properties, e.g., skip

Deblocking Filter

- Similar to what is in VP9
- Changed the order edges are filtered to make hardware easier
- More flexible strength signaling
  - Separate H + V strength for luma
  - Separate C_b and C_r strengths for chroma
  - Can be adjusted on a per-super block basis
- NB: deblocking filter crosses tile boundaries

Constrained Directional Enhancement Filter (CDEF)

- Merge of Daala's directional deringing filter (DERING) and Thor's constrained lowpass filter (CLPF)
  - Both encoder and decoder search for the direction that best matches
  - Primary filter run along direction, and secondary conditional replacement filter run orthogonally
  - Strength is signaled in the bitstream
- Results exceed both DERING and CLPF alone, as well as applying DERING + CLPF sequentially



Loop Restoration

- Enhanced and simplified loop filters from VP10
- Two filter choices per superblock
  - Separable Wiener filter with explicitly coded coefficients
  - Self-guided filter
- Runs in a separate pass after CDEF
  - Showed best metrics of any approach tested
  - Uses deblocking filter output outside of superblock boundaries to minimize line buffers

Spatial and Temporal Scalability

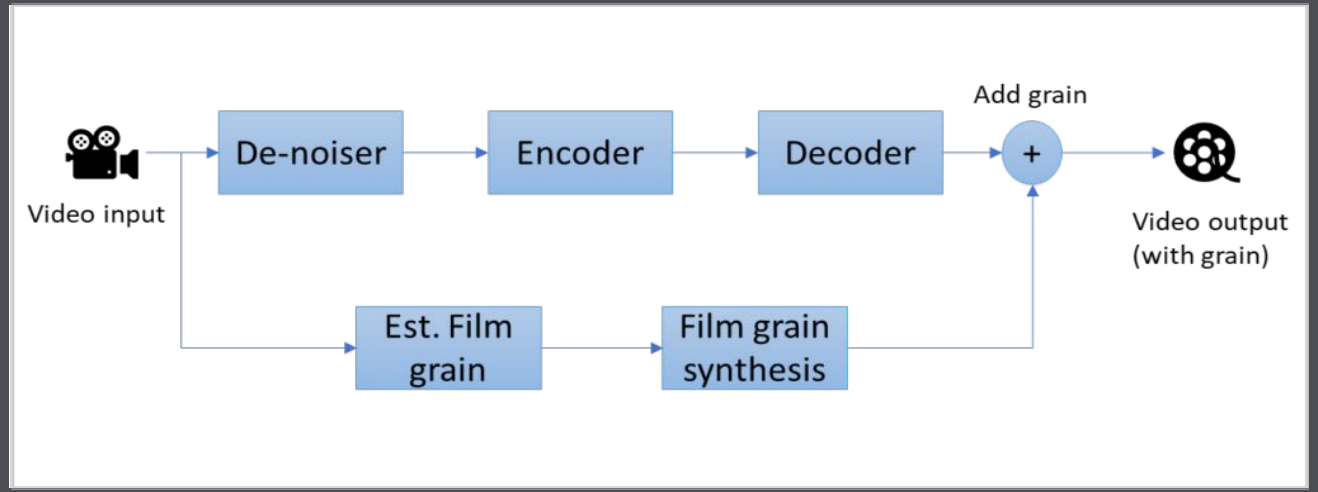
- Each frame can have a `spatial_id` and a `temporal_id`
  - When `spatial_id` = 0 and `temporal_id` = 0 it is called a base layer
  - When `spatial_id` > 0 and `temporal_id` > 0 it is called an enhancement layer
- Idea is that decoder will simply display the frames from the highest layer
  - Higher layer frames can reference lower layer frames
- Designed to be used by a special “Selective Forwarding Unit” server that hands out the appropriate scalable layer to a client

Frame Super-Resolution

- Not actually super-resolution
- Instead
  - Code at reduced resolution
    - Run deblocking filter and CDEF, but not Loop Restoration filter
  - Upsample with simple upscaler
  - Run Loop Restoration filter at full resolution
- Only horizontal resolution reduction allowed
  - Simplifies hardware (no new line buffers)
- Allows for gradual bitrate scaling

Film Grain Synthesis

- Grain parameters signaled per frame
- Synthesized film grain applied after decoding (not in loop)
- Could be applied using GLSL + PRNG based texture



AOM Members / Hardware

Designed for Hardware Implementations

Hardware members involved from the very beginning

Feedback incorporated into a number of tools

- Per symbol probability adaptation
- Smaller multipliers in entropy coder
- Single pass bitstream writing
- Fewer line buffers in CDEF and LR
- Only allow horizontal scaling for super-resolution

AOM Members / Real-Time Conferencing

Designed for Low-Latency

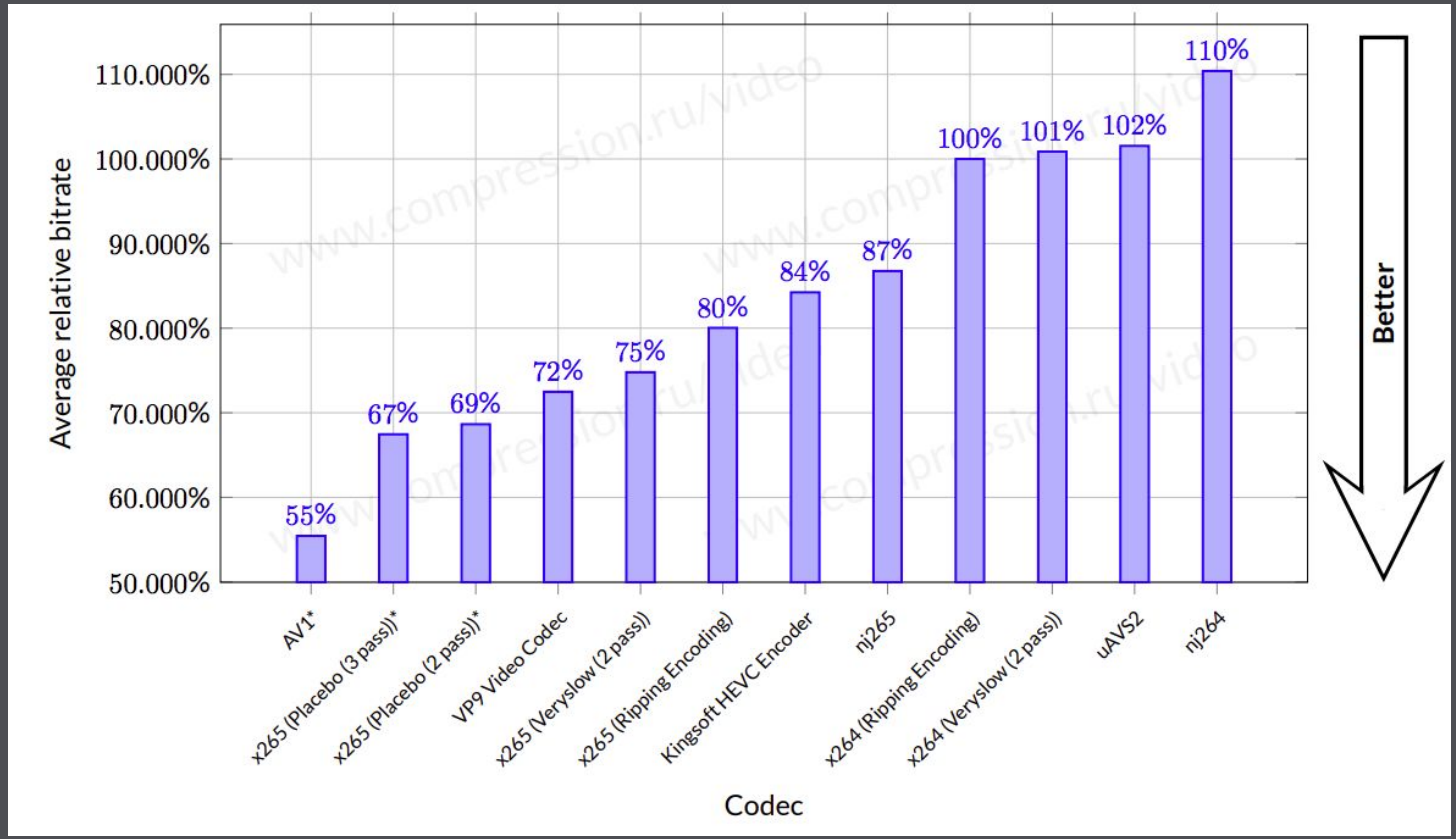
Per symbol adaptation replaces symbol counts in VP9
Can write bitstream with subframe latency
Removed signaling from frame header that forced whole frame buffering

Designed for Broadcasters?

- Decoder rate model
  - Guarantee buffer size
  - Limit the use of alt-ref's to ensure decodability
  - Verifiable (See Annex E of the spec document)

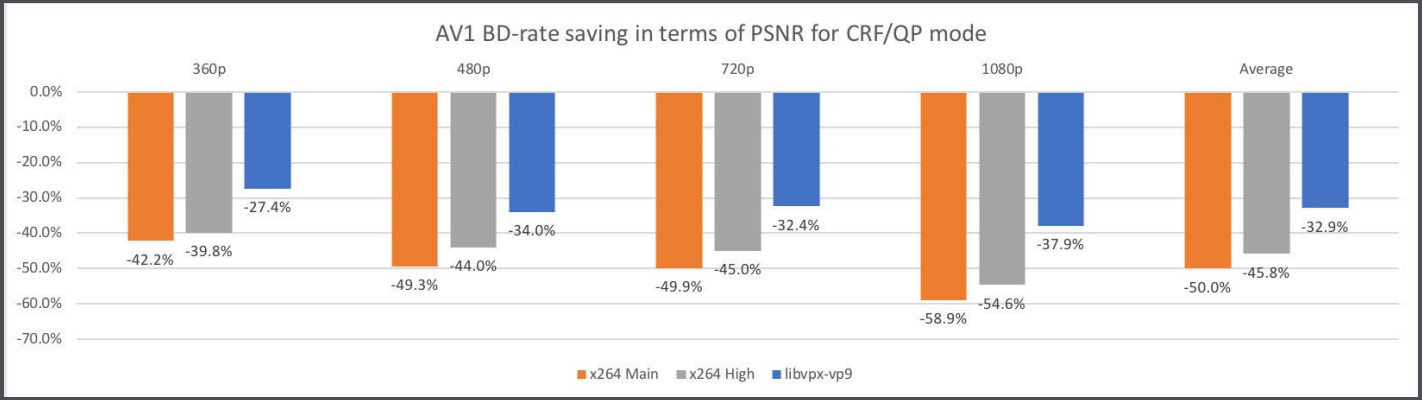
- Support for AV1 coming to hardware
  - Smart TV's will want to play Netflix, Hulu, YouTube, etc.
- Start with AV1 in the broadcasting stack
  - Can leverage industry investment in hardware, software, tooling, etc.
  - Easier to expand into streaming market

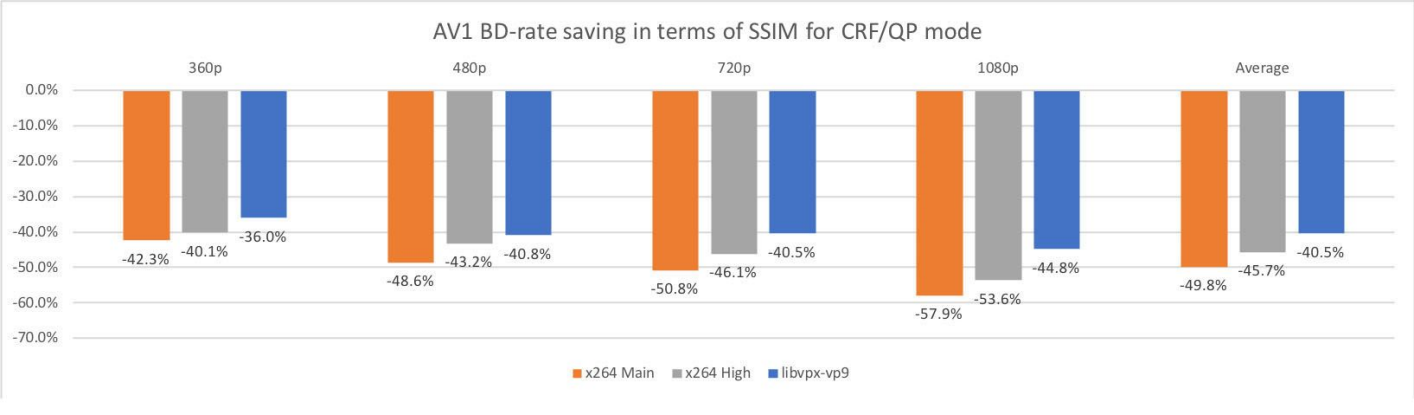
Moscow State University (SSIM - June 2017)



http://www.compression.ru/video/codec_comparison/hevc_2017/MSU_HEVC_comparison_2017_P5_HQ_encoders.pdf

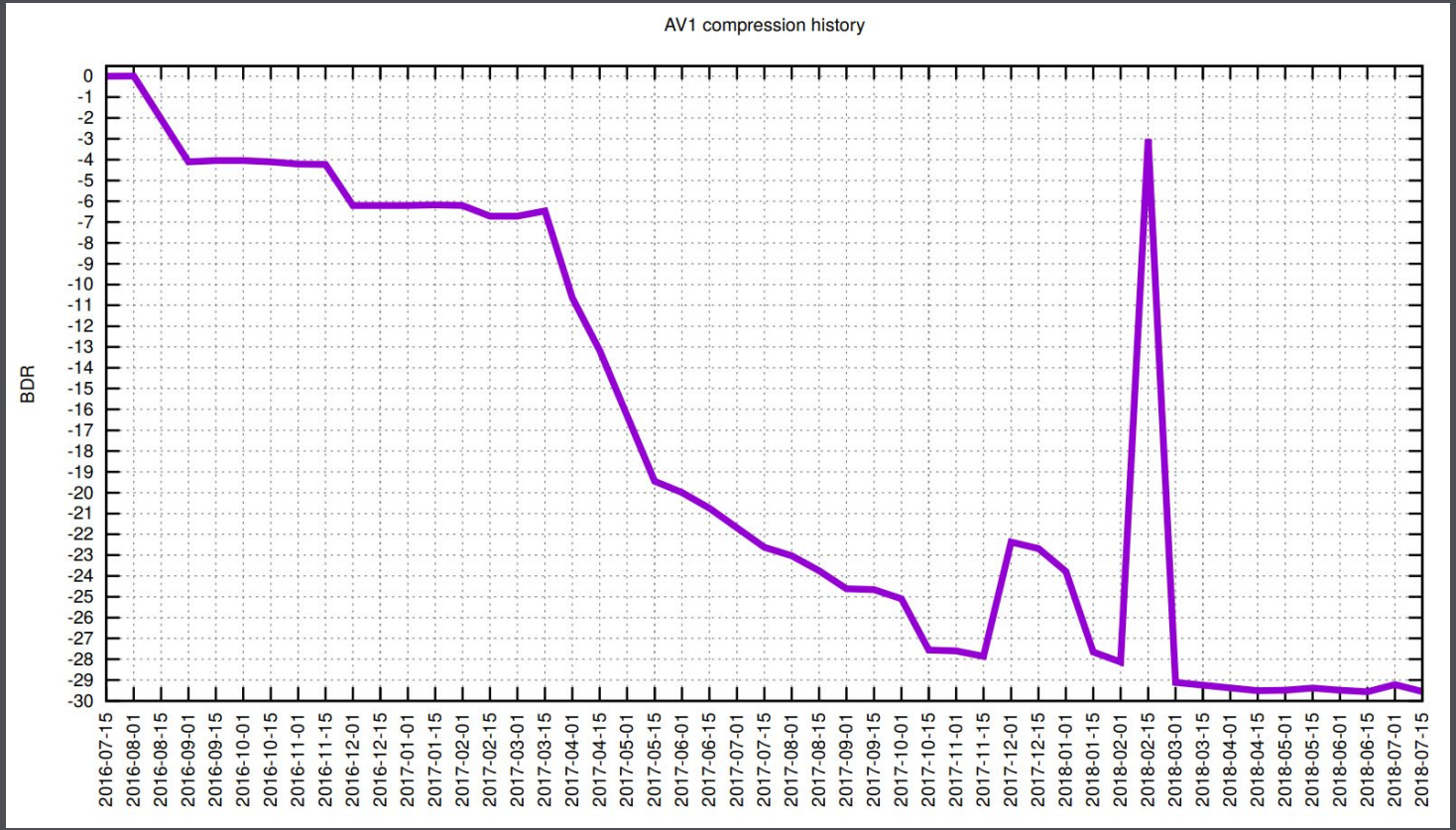
Facebook Study (April 2018)



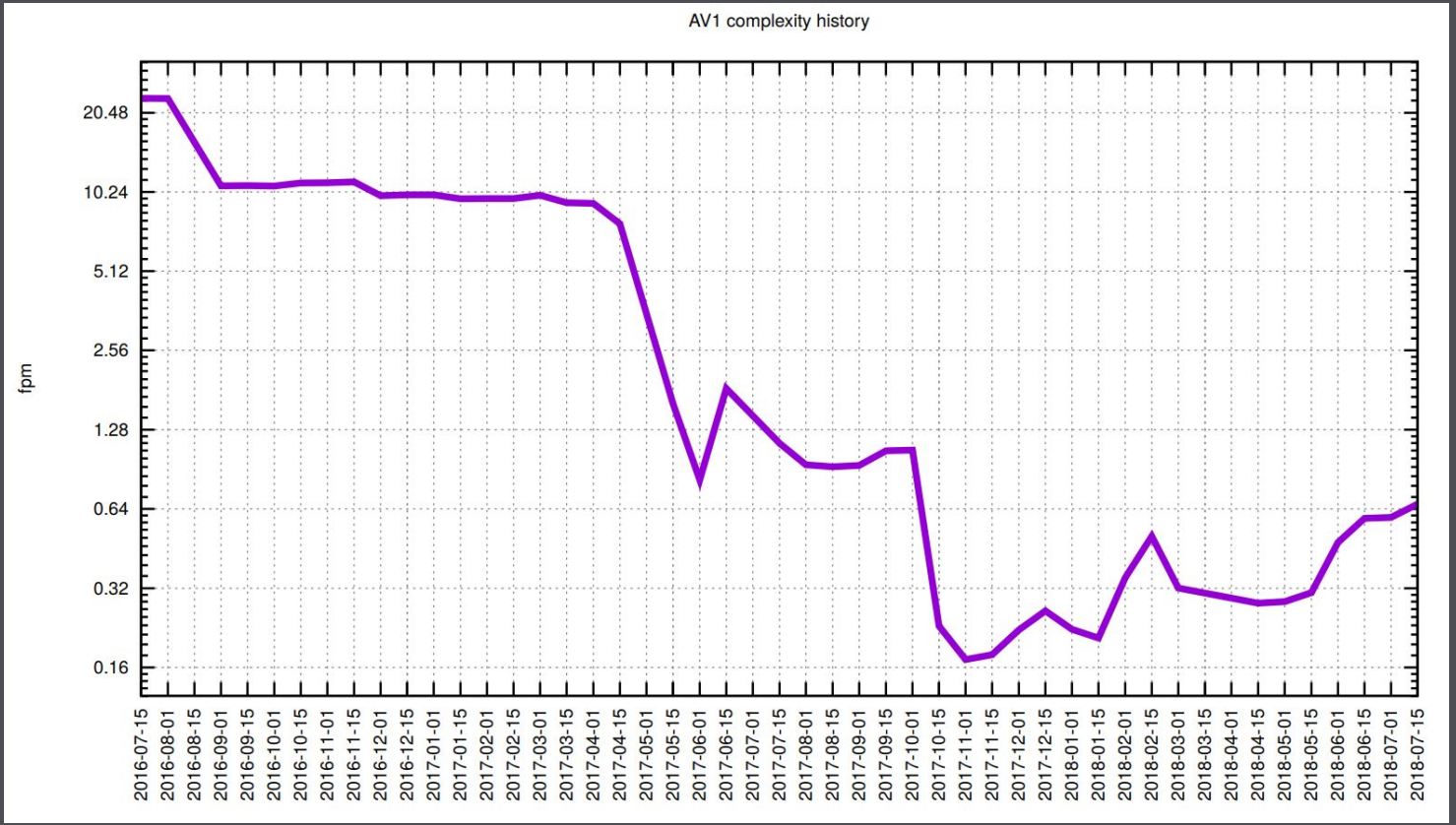


https://code.fb.com/video-engineering/av1-beats-x264-and-libvpx-vp9-in-practical-use-case/

AV1 Compression History



AV1 Complexity History



moz://a

Questions?