AV1: The Quest is Nearly Complete

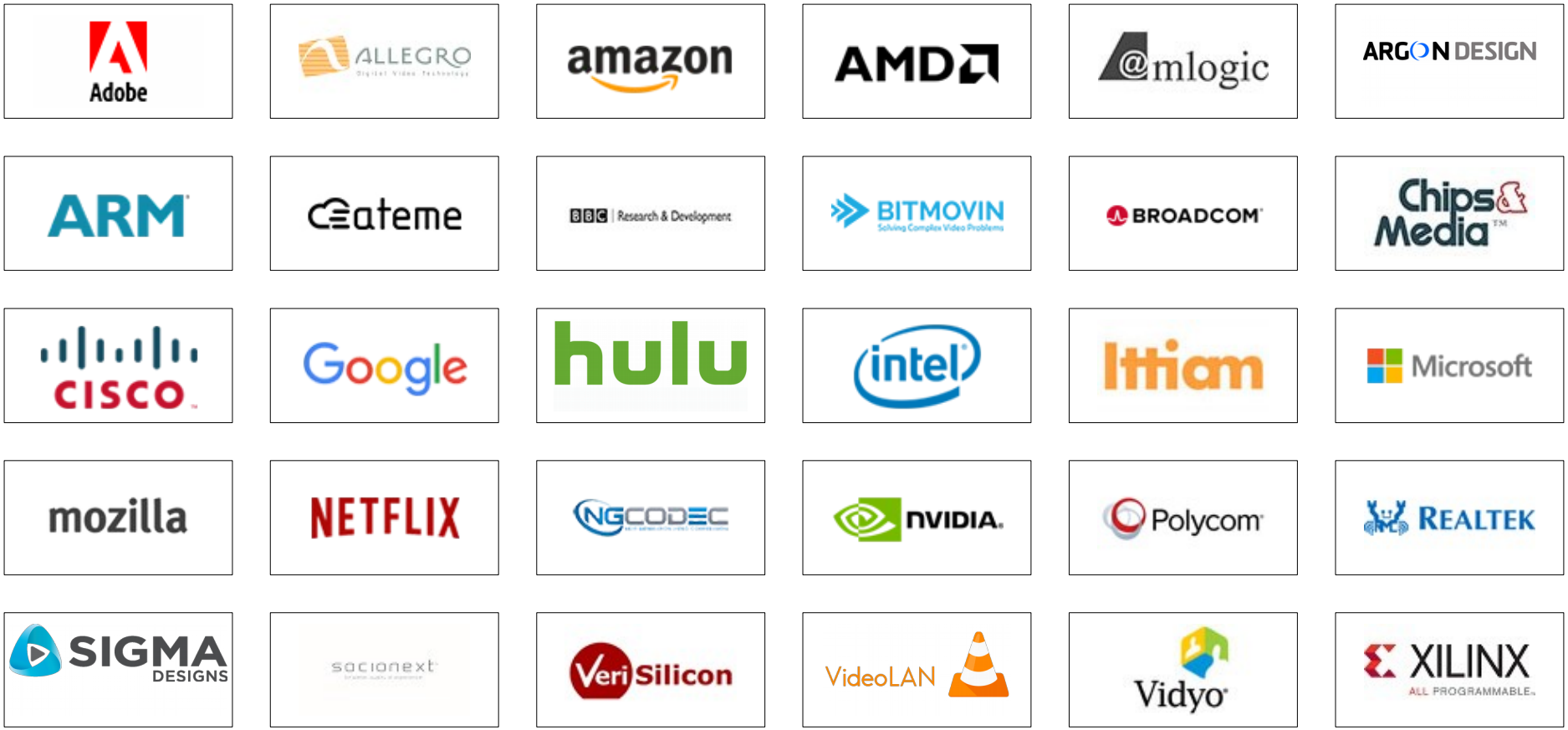
Thomas Daede
tdaede@mozilla.com
October 22, 2017

slides: https://people.xiph.org/~tdaede/gstreamer_av1_2017.pdf

Mozilla & The Xiph.Org Foundation

Who are we?

- Joint effort by lots of companies to develop a royalty-free video codec for the web



Video format	Licensor	Codec Royalties	Codec Royalty Exemptions	Codec Royalty Annual Cap	Content Distribution Fee
HEVC	MPEG LA	■ US$0.20 per unit	■ First 100k units each year[157]	■ US$25 million	■ US$0
	HEVC Advance	**Region 1**: ■ US$0.40 (mobile) ■ US$0.80 (PC) ■ US$1.20 (TV) **Region 2**: ■ US$0.20 (mobile) ■ US$0.40 (PC) ■ US$0.60 (TV)[158]	■ US$25,000 each year[159] ■ Most software HEVC implementations distributed to consumer devices after first sale[160]	■ US$40 million	**For content that is free to end users**: ■ US$0 **Paid By Title Model**: ■ US$0.025 per paid title **Paid Subscriber Model**: ■ US$0.005 per month per paid subscriber ■ US$2.5 million annual cap per business model ■ US$5 million total annual cap
	Technicolor	tailor-made agreements[48]			■ US$0[48]
	Velos Media[50]				
	others (AT&T, Microsoft, Motorola, Nokia, Cisco, ...)[42][161][162]				

Current Status

- Planning soft bitstream freeze by the end of the month!
- Lots of decisions made
- Some tools still have work remaining
- IPR analysis ongoing

How AV1 Works

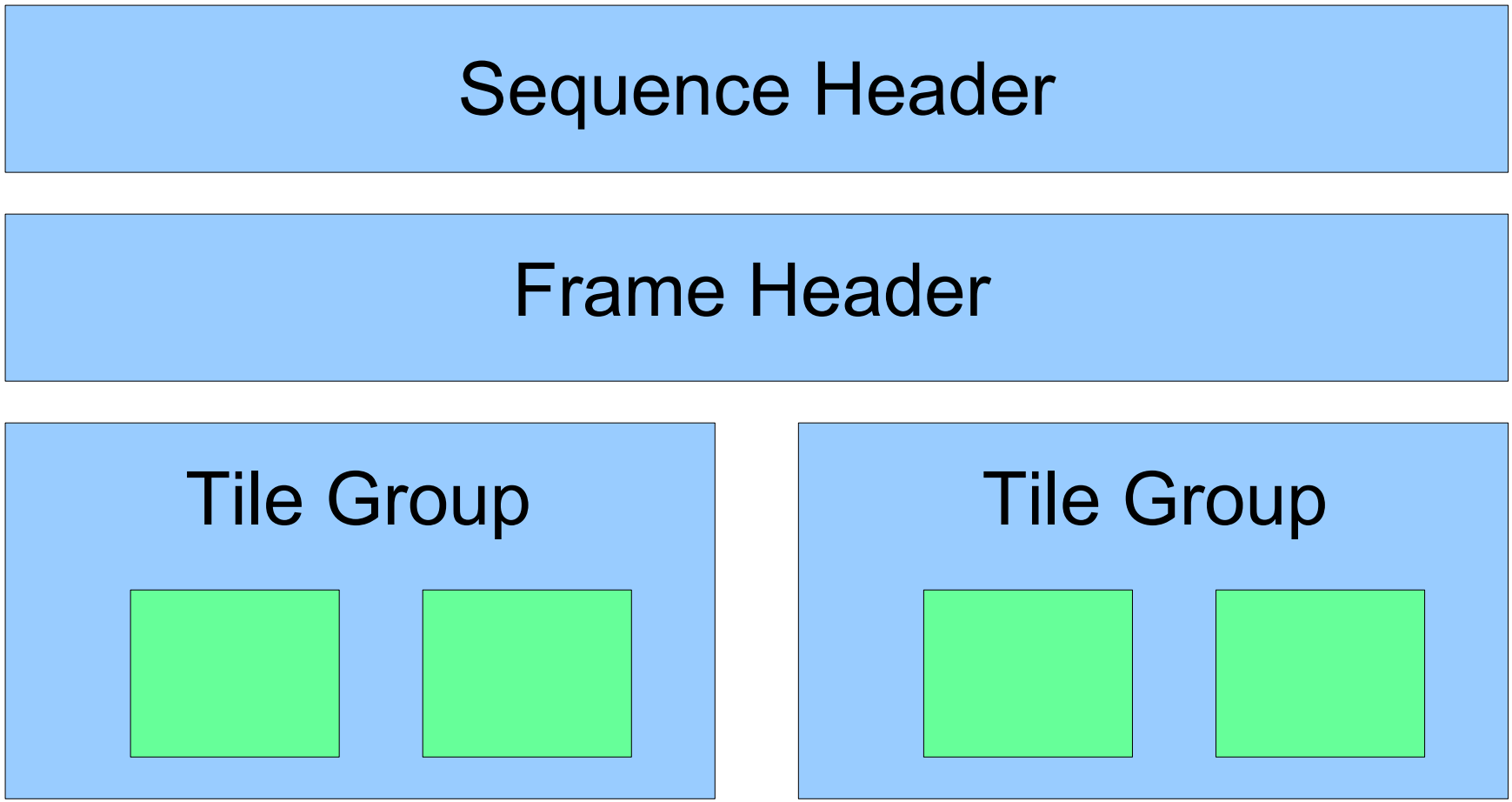
Lots of stuff!

- New high-level syntax
  - Easily parseable sequence header, frame header, tile header, etc.
- New adaptive multisymbol entropy coding
- More block sizes
  - Prediction blocks from 128x128 down to 4x4
    - Includes rectangular blocks with 1:2 and 2:1 (4x8, 8x4, etc.) as well as 1:4 and 4:1 ratios (4x16, 16x4, etc.)
  - Transforms from 32x32 down to 4x4
    - Includes 1:2 and 2:1 rectangular transforms (4x8, 8x4, …)
- More transform types
- More references
  - Up to 7 per frame (out of a store of 8)
- More prediction modes
  - Both intra and inter
- More in-loop filtering

What I'll cover

- Some selected new features
- Containers, ecosystem, etc
- Tooling

High-level Syntax



High-level Syntax

- Assists in easy parsing of bitstream
- Used to define packing into containers
  - Matroska & WebM
  - ISOBMFF (MP4, DASH, HLS)
    - gstreamer's hlssink doesn't support this yet
  - RTP (WebRTC)
- Other containers can also be supported

Colors and HDR

- Colorspace, color matrix, transfer function can now be encoded directly in the bitstream
  - Chroma siting and levels too

Intra Prediction

Intra Prediction

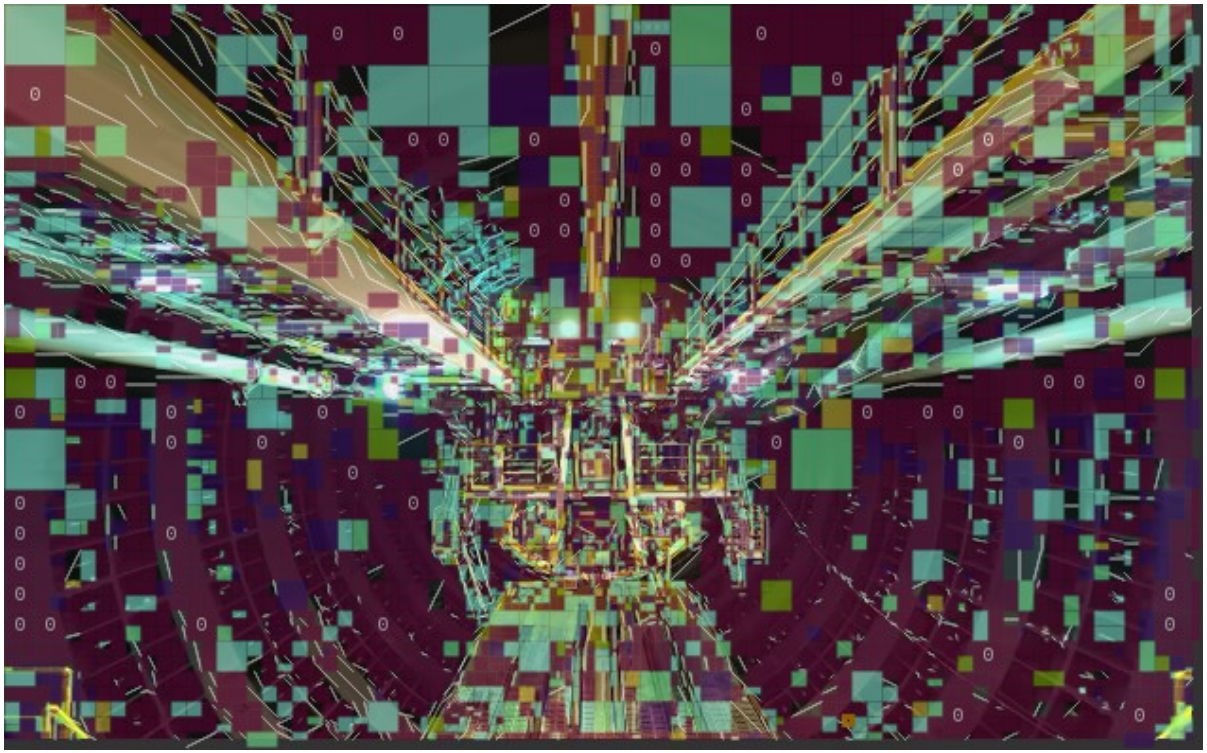
UV Mode Selection Example (https://goo.gl/6tKaB8)

CFL_PRED 17%

DC_PRED 44.36%

TM_PRED 7.98%

SMOOTH_PRED 4.85%



Ohashi0806shield.y4m
QP = 55

Intra Prediction Modes

- More directional modes
  - 8 main directions plus delta for up to 56 directions
  - Not all available at smaller block sizes
- Smooth 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

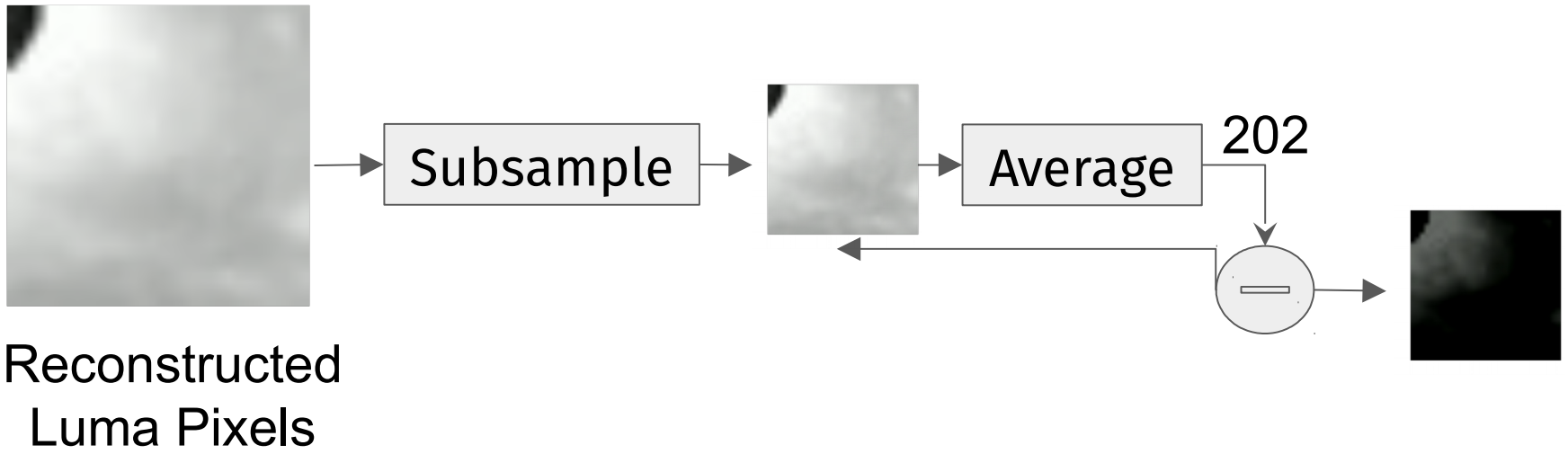
Other Intra Prediction Enhancements

- Blend neighbor pixels before prediction
  - Strength depends on prediction angle (relative to border orientation) and block size
- Edge extension
  - If pixels from one neighboring block unavailable, extend from an adjacent neighboring block edge
- Chroma from Luma

Chroma from Luma

Chroma from Luma

- Step 1: Compute AC Contribution

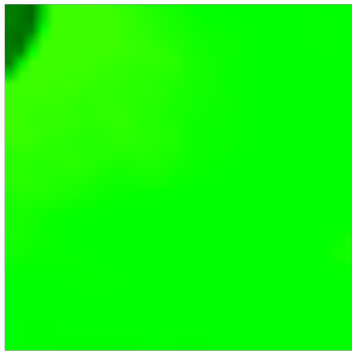


Chroma from Luma

- Step 2: Scale Chroma Planes

$\alpha_{Cb} = -0.25$

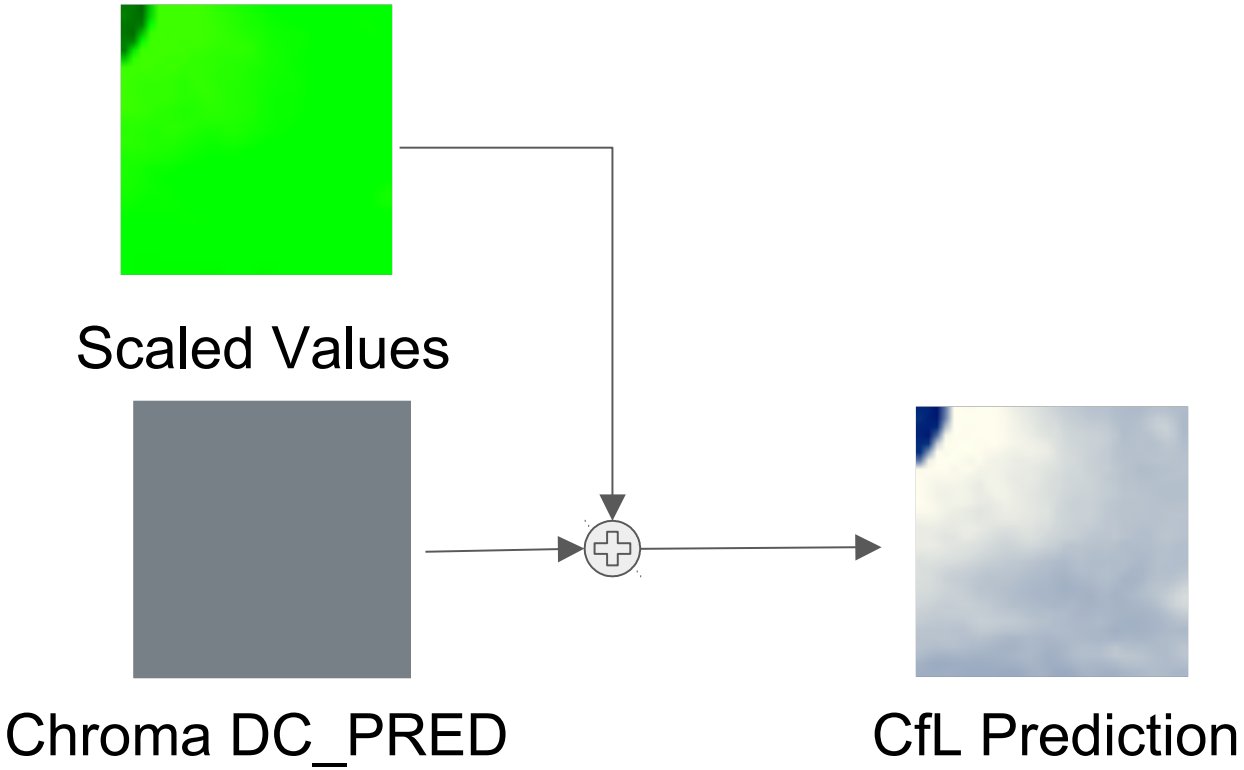
$\alpha_{Cr} = 0.125$



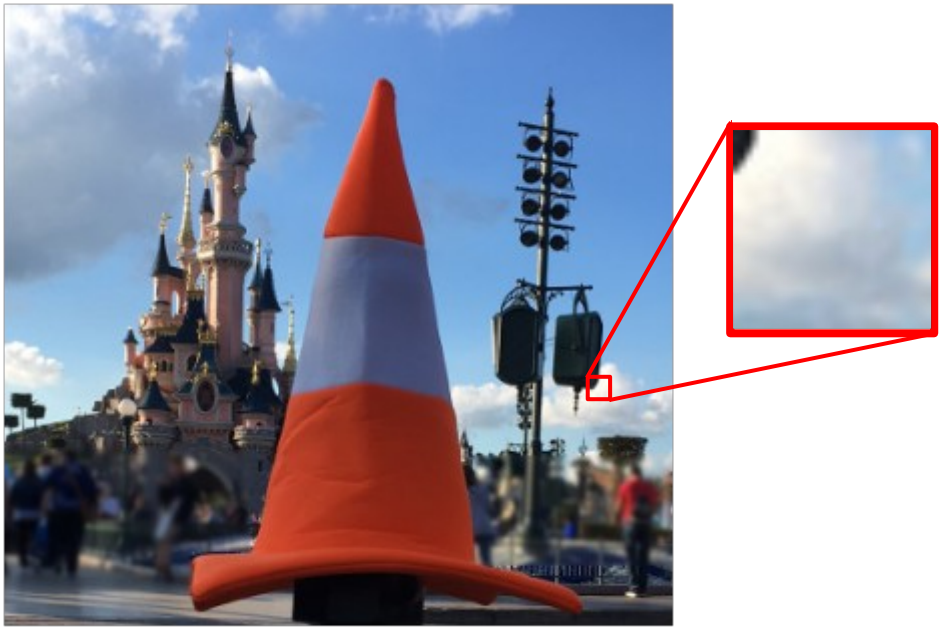
Scaled Values

Chroma from Luma

- Step 3: Add Chroma DC_PRED



CFL in Action



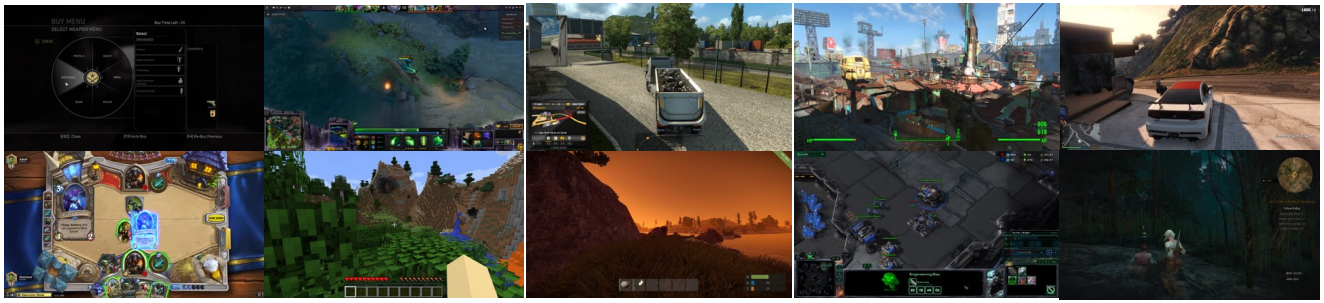
Chroma DC_PRED

CFL_PRED
Scaling factors (-0.25, 0.125)

Awesome for Gaming

Video	PSNR	PSNR HVS	SSIM	CIEDE 2000	PSNR Cb	PSNR Cr
Average	**-1.01**	**-0.93**	**-0.90**	**-5.74**	**-15.58**	**-9.96**
CSGO_0_120f.y4m	-0.70	-0.34	-0.01	-1.12	-12.00	27.02
CSGO_10_120f.y4m	-1.63	-1.33	-1.42	-0.35	-7.43	0.74
DOTA2_0_120f.y4m	-0.46	-0.46	-0.42	-2.49	-5.29	-6.42
DOTA2_10_120f.y4m	-0.69	-0.74	-0.53	-3.14	-5.99	-5.78
EuroTruckSimulator2_0_120f.y4m	-0.17	-0.24	-0.07	-2.73	-11.11	-10.13
EuroTruckSimulator2_10_120f.y4m	-0.40	-0.38	-0.43	-3.45	-12.05	-10.96
Fallout4_0_120f.y4m	-0.44	-0.29	-0.21	-4.33	-10.17	-7.49
Fallout4_10_120f.y4m	-0.45	-0.35	-0.31	-3.57	-8.33	-5.43
GTAV_0_120f.y4m	-1.11	-1.11	-1.01	-5.88	-15.39	-5.57
GTAV_10_120f.y4m	-0.84	-0.94	-0.88	-8.34	-41.20	-9.69
Hearthstone_0_120f.y4m	-1.30	-1.34	-1.27	-2.64	-4.40	-3.60
Hearthstone_10_120f.y4m	-0.67	-0.73	-0.61	-2.43	-4.57	-4.08
MINECRAFT_0_120f.y4m	-4.07	-3.50	-3.92	-19.68	-37.90	-29.09
MINECRAFT_10_120f.y4m	-3.76	-3.12	-3.68	-20.69	-31.44	-25.54
RUST_0_120f.y4m	-0.36	-0.36	-0.28	-6.31	-25.17	-21.35
RUST_10_120f.y4m	0.25	0.25	0.31	-6.34	-23.90	-27.68
STARCRAFT_0_120f.y4m	-0.85	-0.94	-0.87	-6.05	-12.58	-16.99
STARCRAFT_10_120f.y4m	-1.41	-1.43	-1.38	-4.15	-6.18	-6.21
WITCHER3_0_120f.y4m	-0.21	-0.25	-0.20	-5.41	-20.92	-21.06

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



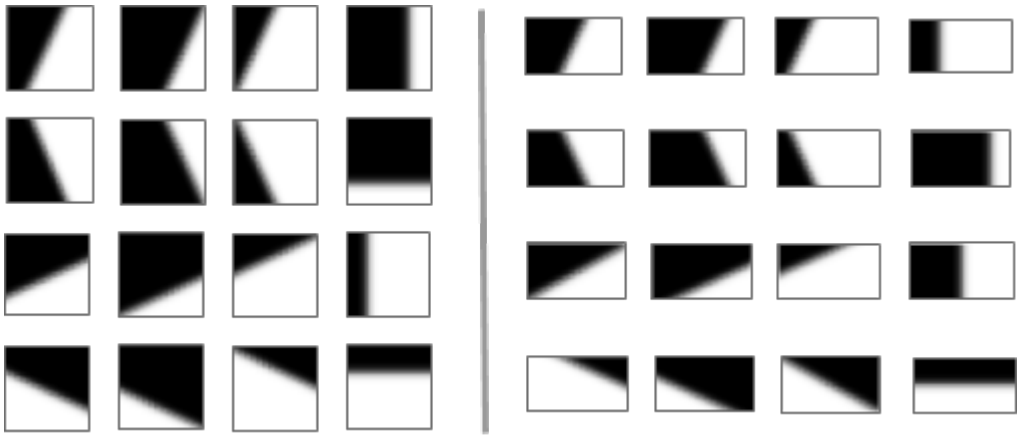
Inter Prediction

Motion Vector Coding

- Each frame has a list of 7 previous frames to reference (out of a pool of 8)
- Construct list of top 4 MVs for given reference/reference pair from neighboring areas

Compound Modes

- Inter-Inter Compound Segment
  - Pixel weight depends on difference between prediction pixels
- Inter-Intra gradual weighting
  - Smoothly blends from inter to intra prediction
  - Wedge codebook (Inter-Inter or Inter-Intra)



Square Codebook

Rectangular Codebook

Global Motion

- Defines up to a 6-parameter affine model for the whole frame (translation, rotation, 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

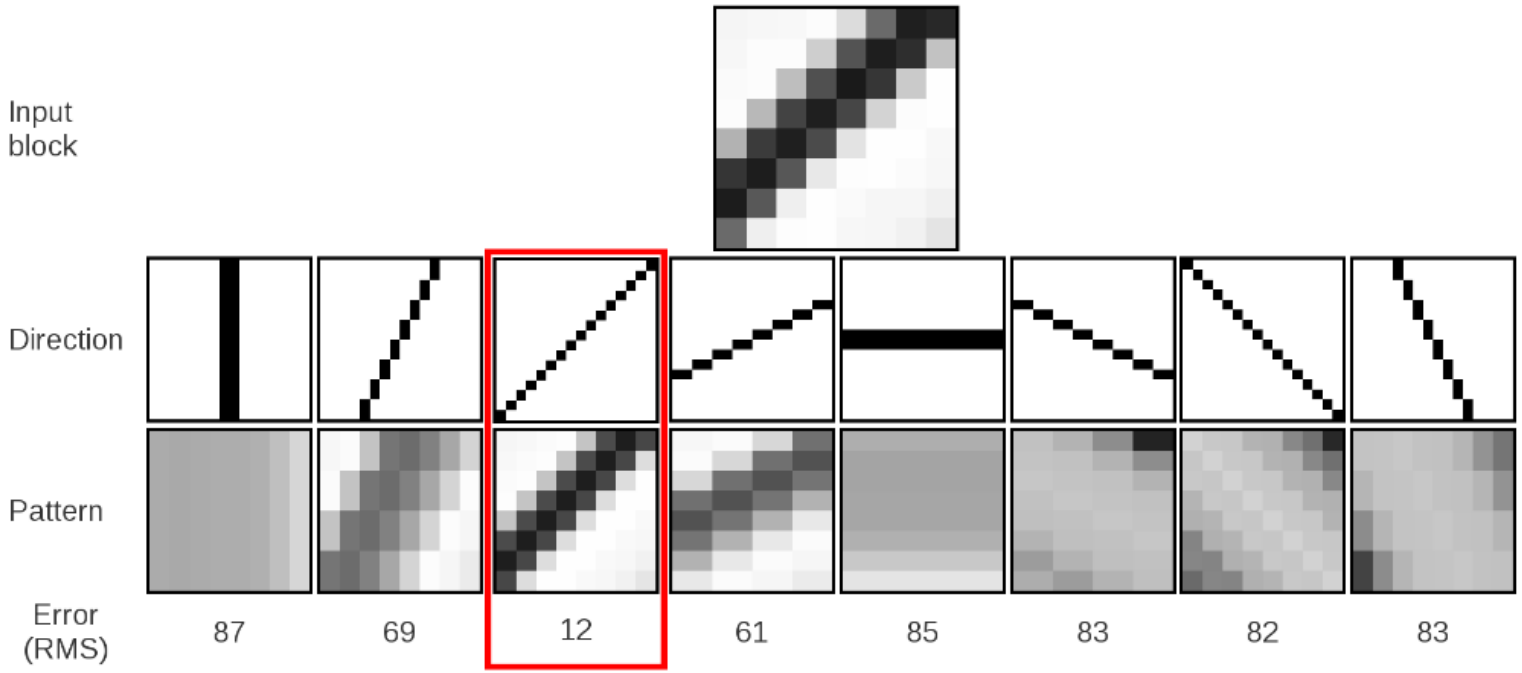
Warped Motion

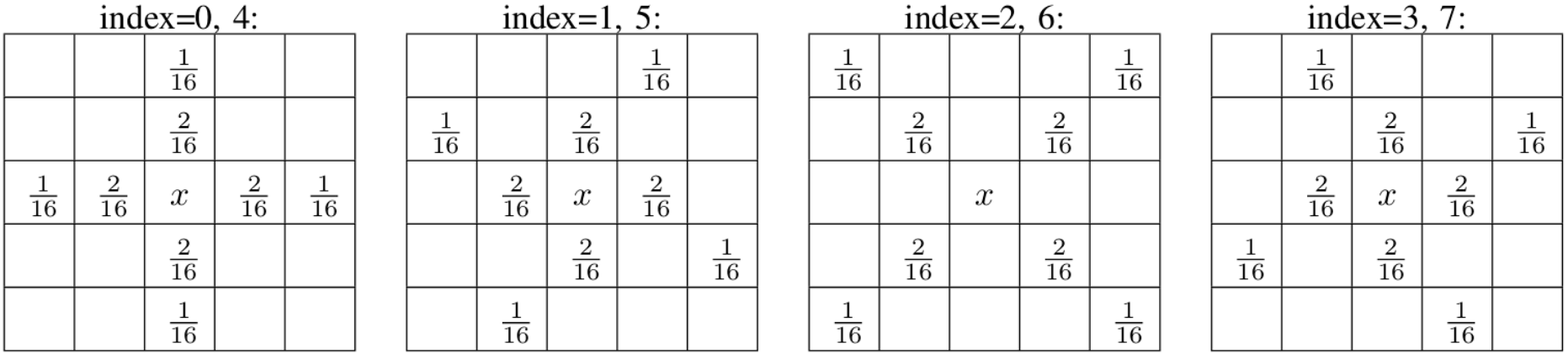
- Use neighboring blocks to define same motion model within a block
  - Decomposed into two shears with limited range
    - Similar complexity to subpel interpolation

Loop Filtering

Constrained Directional Enhancement Filter

- Combines Daala's directional deringing filter and Thor's Constrained Low-Pass Filter (CLPF)





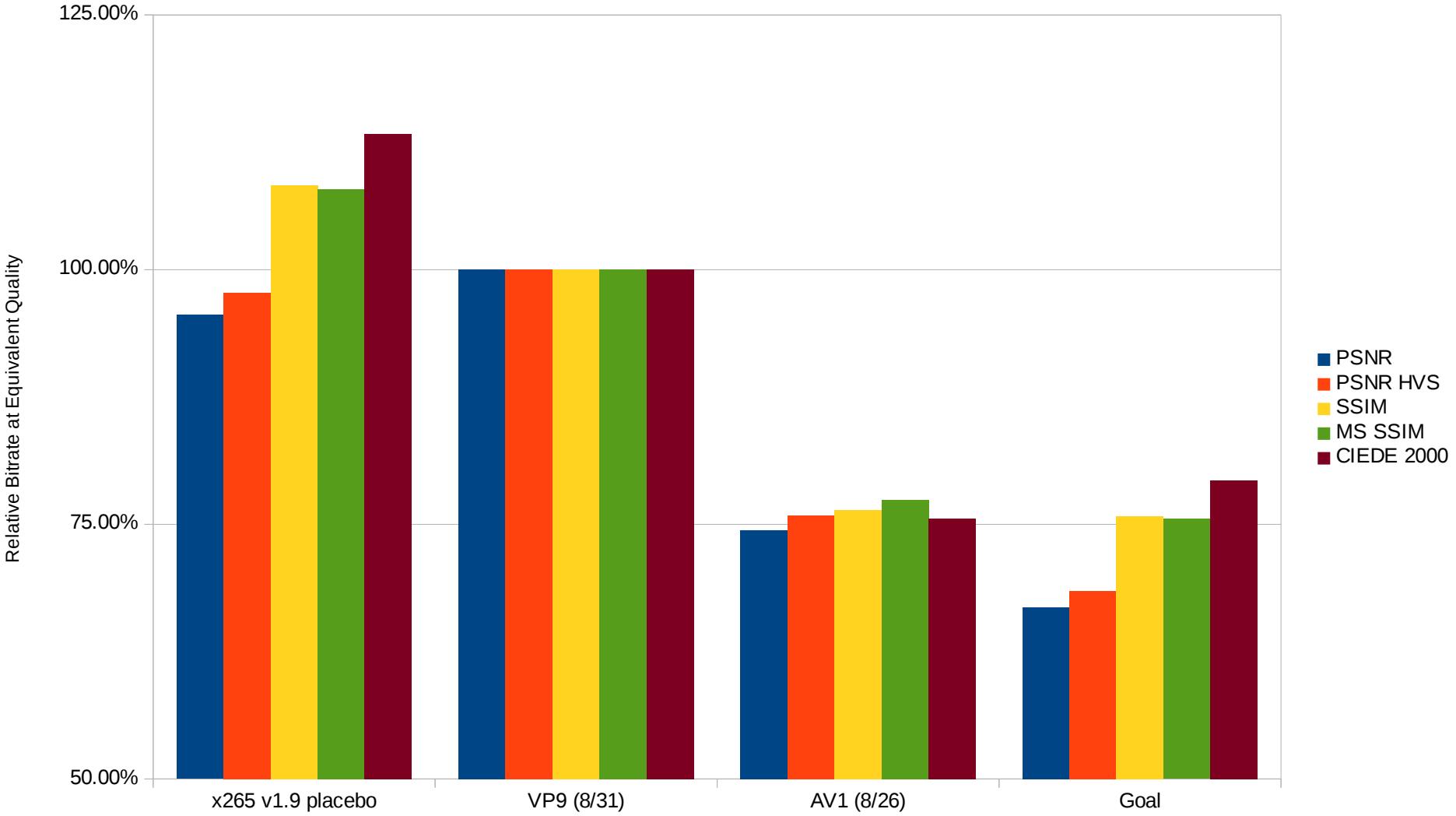
- Single-pass design
  - Both filters applied simultaneously
  - Fewer line buffers in hardware compared to a simple combination

Loop Restoration

- 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

Metrics

Metrics



Follow link for AWCY details

Complexity

- AV1 gets most of its compression gains by adding more tools and more options
  - More partition and transform sizes
  - More prediction modes with more parameters
  - More transform types
- Searching them all in the encoder is slow
  - About 200x slower than VP9 right now
- Decoder also slower, but not as much
  - Hard to give precise numbers in unoptimized state

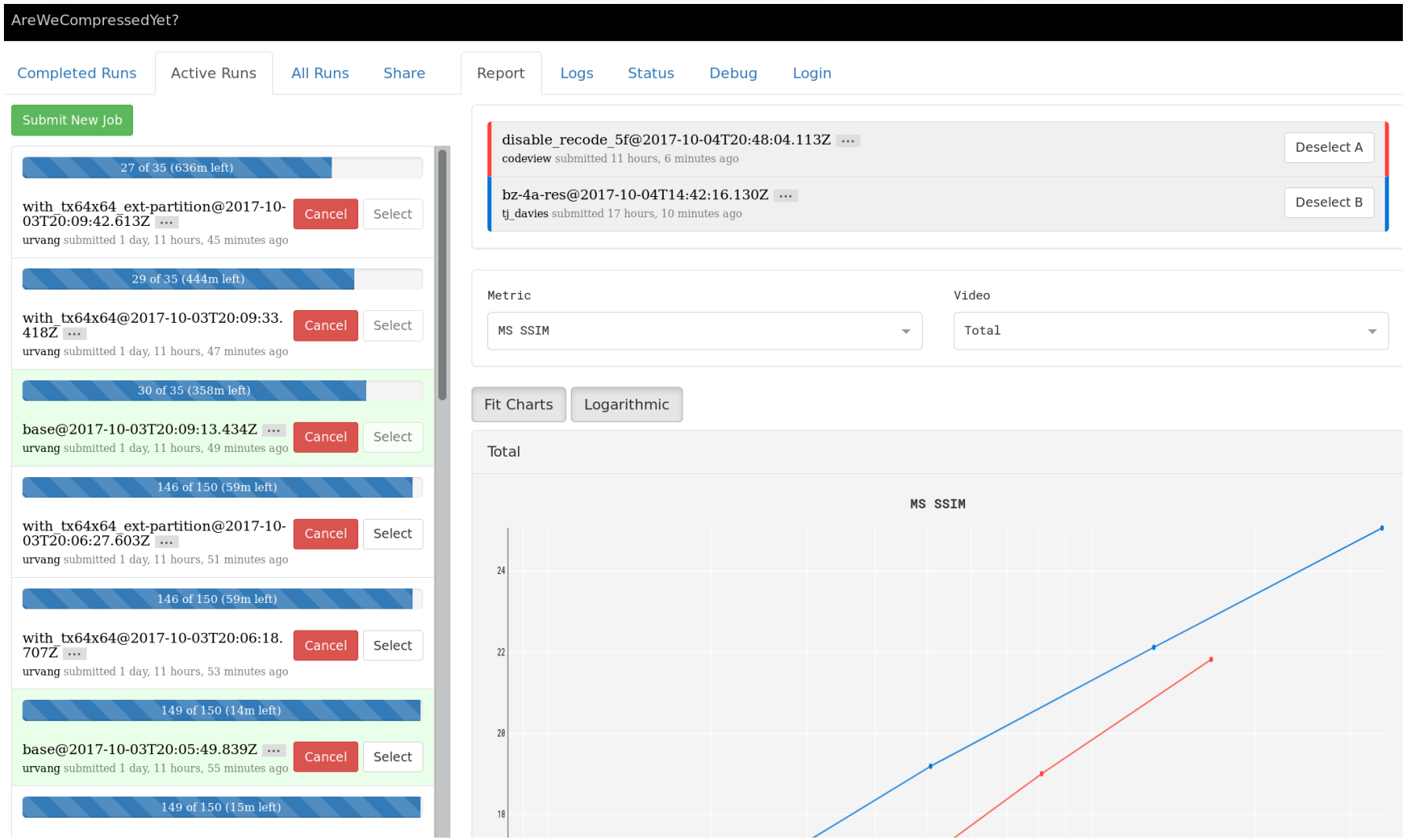
Future encoder improvements

- Quantization matrices
- Better delta-QP, segments
- New experiments (dist_8x8) add better distortion functions for RDO
  - daala
  - cdef

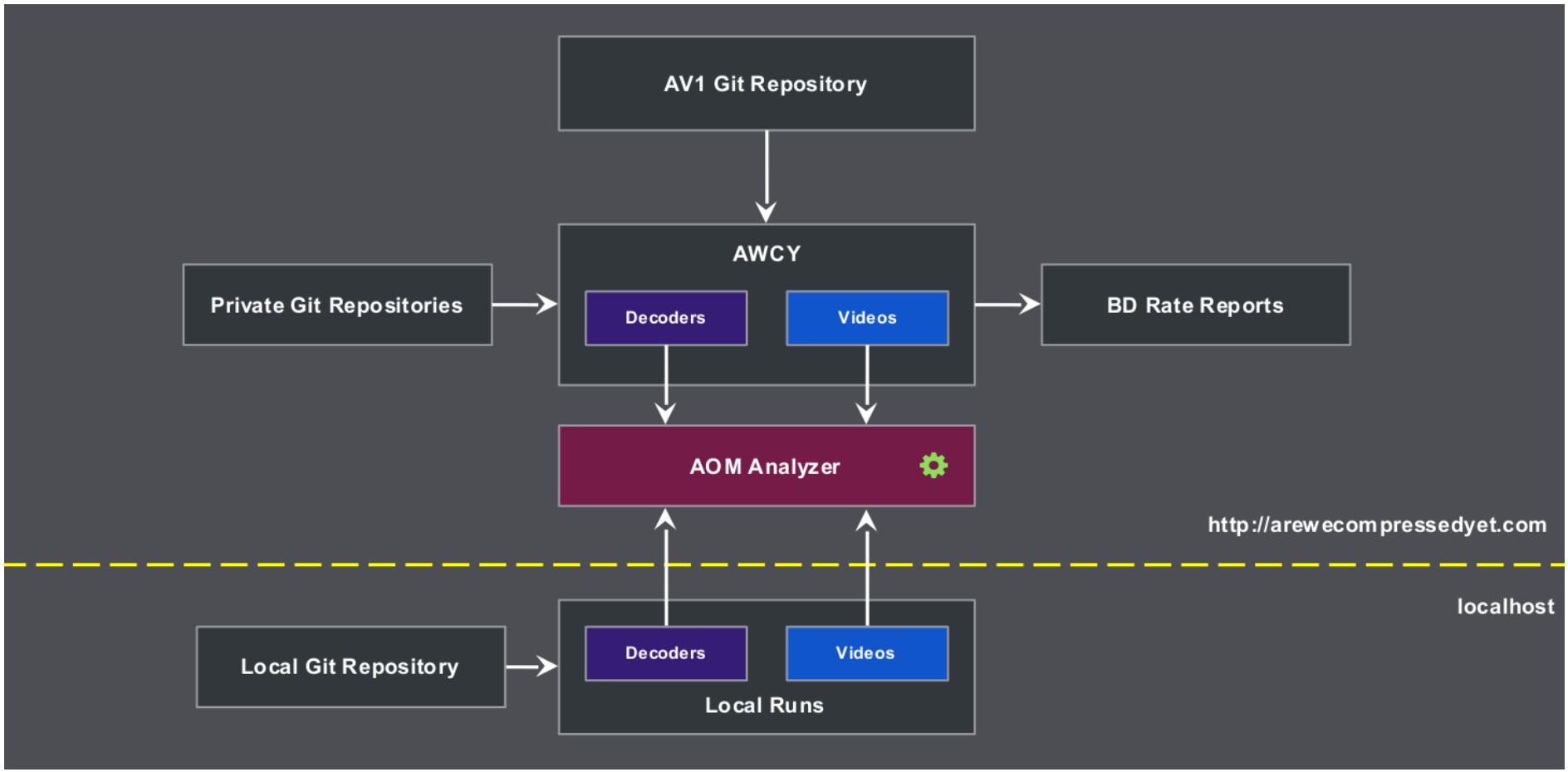
Tools (the utility kind)

- AreWeCompressedYet
- AOM Analyzer
- Subjective viewing

AreWeCompressedYet

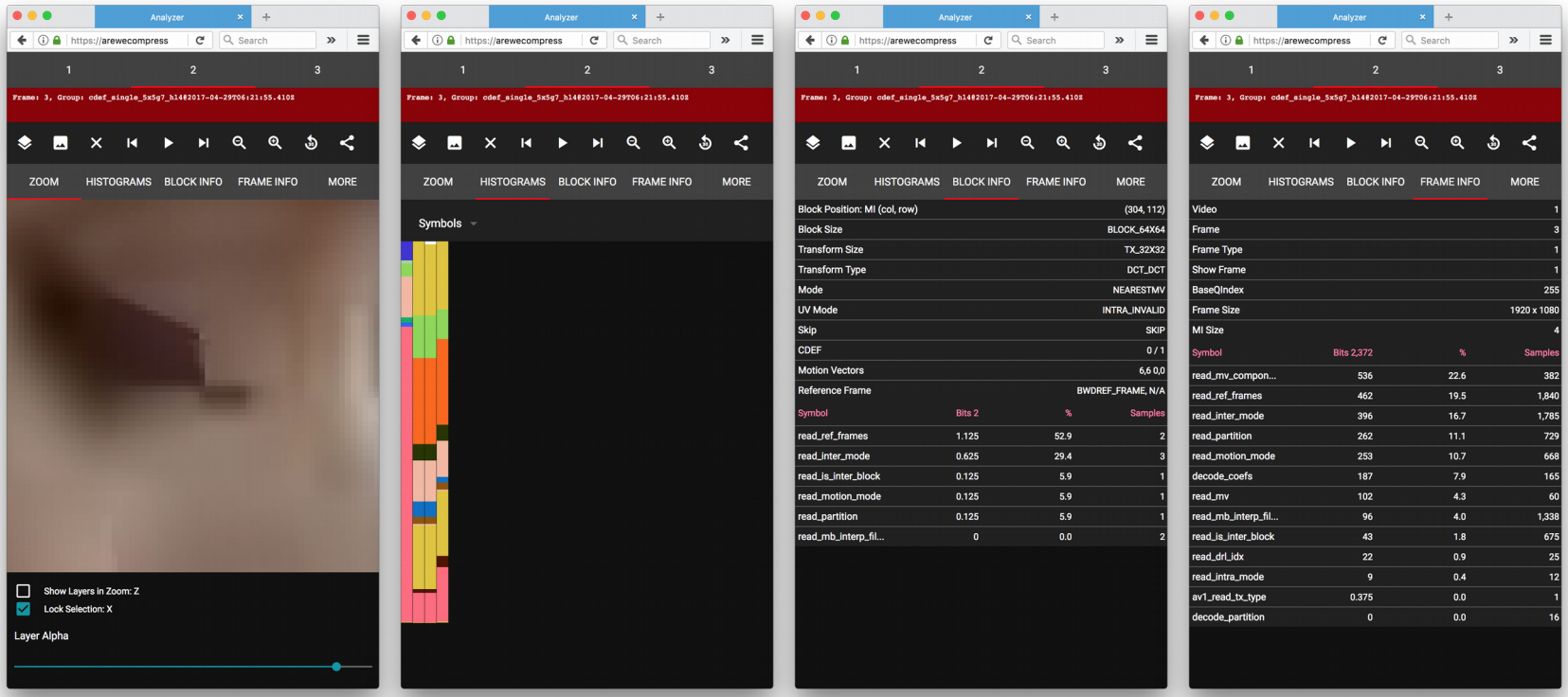


AOM Analyzer

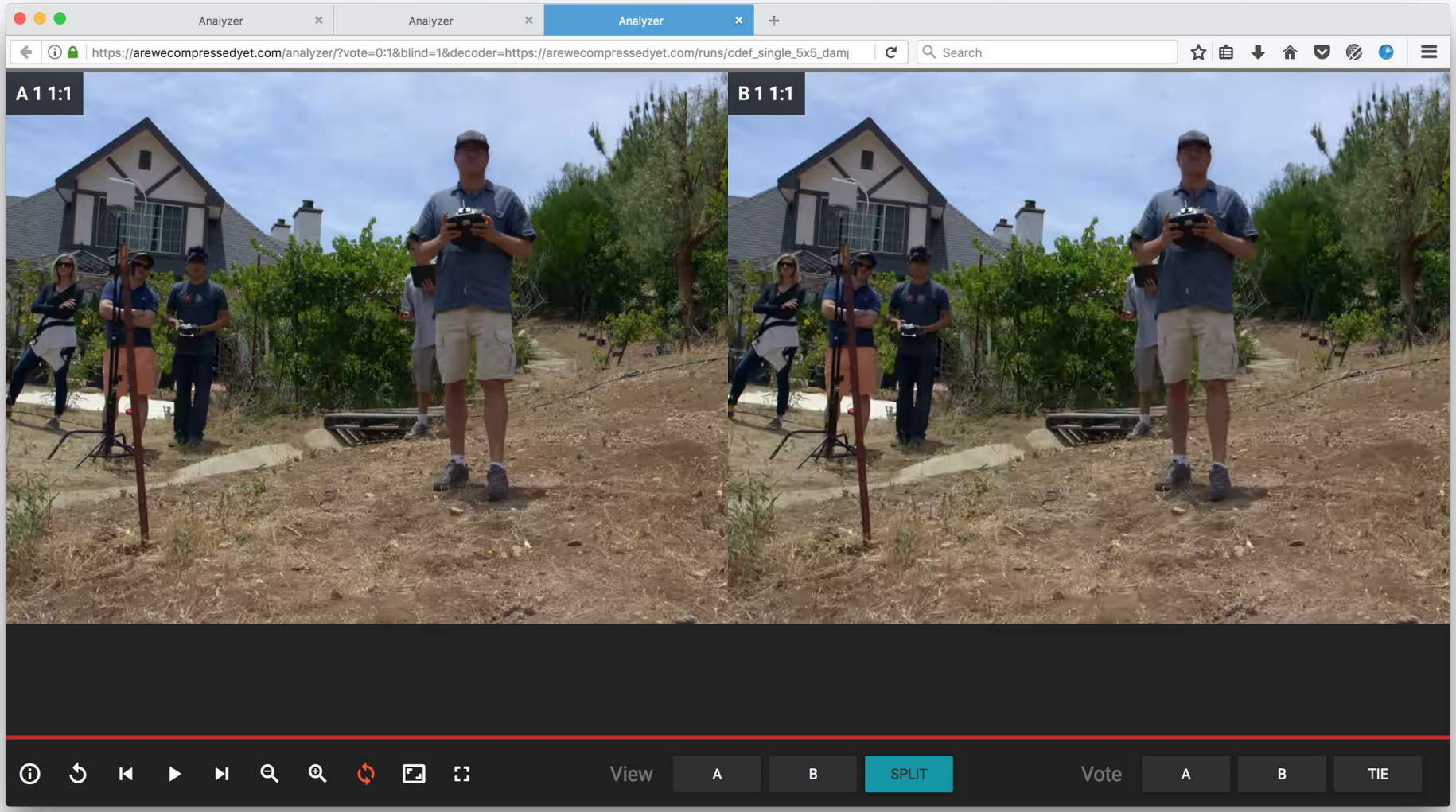


AOM Analyzer

Tabs



Subjective testing



Implementations

- libaom
  - Reference implementation, similar API to libvpx but not compatible
    - https://aomedia.googlesource.com/aom/
- rav1e
  - Encoder only, very fast but very low quality
  - gstreamer-rs bindings?
    - https://github.com/tdaede/rav1e

Code

- Gerrit code review
  - https://aomedia-review.googlesource.com/
- AOM Analyzer
  - https://github.com/mbebenita/aomanalyzer
- AreWeCompressedYet
  - https://github.com/tdaede/awcy
- Specification
  - https://aomedia.googlesource.com/av1-spec/

Demo

https://people.xiph.org/~tdaede/demuxed.webm

Questions?