AOM/AV1:
Next Generation Open Video Codec

Zoe Liu (zoeliu@google.com)
Google Inc.

Google

Who are we?

Chrome Media

Our mission: Open, free media technologies for the web

Video: **VP8**, **VP9**, **AV1** (in development by AOMedia)

Still image: **WebP**

Audio: **Opus**

Geometry compression: **Draco**

Open Source Video Codecs

Why open codecs?

Video for the Web

Video for Broadcast/Disc

Why open codecs?

Fundamental to the Web: Royalty-free, open-source is at need

- Technology carrying royalties creates a barrier to smaller players in the market
- Open source provides benefits of security

Faster innovation: break out of standard process

What has VP9 achieved?

Royalty free

50% bitrate reduction vs. very best H.264

Advanced profiles support 10-/12-bit and higher color spaces (4:4:4 / 4:2:2)

What has VP9 achieved?

VP9 supported in billions of end-points

Support in Chrome, Firefox, Edge, Opera, Android 4.4 (Kk) forward

Broad support for Profile 0 (8-bit) on living room devices

Wide support for Profile 2 (10-bit) on living room devices

Youtube + VP9

2.5 billion hours of VP9 video served in the first year

2+ billion views per day in VP9

Latency is reduced by 15% everywhere

Buffering is improved by 25%/100% in developed/emerging market

Youtube HD delivery is increased by 25% in emerging market

Youtube + VP9

VP9

Faster

VP9 has reduced latency by 15% everywhere

Smoother

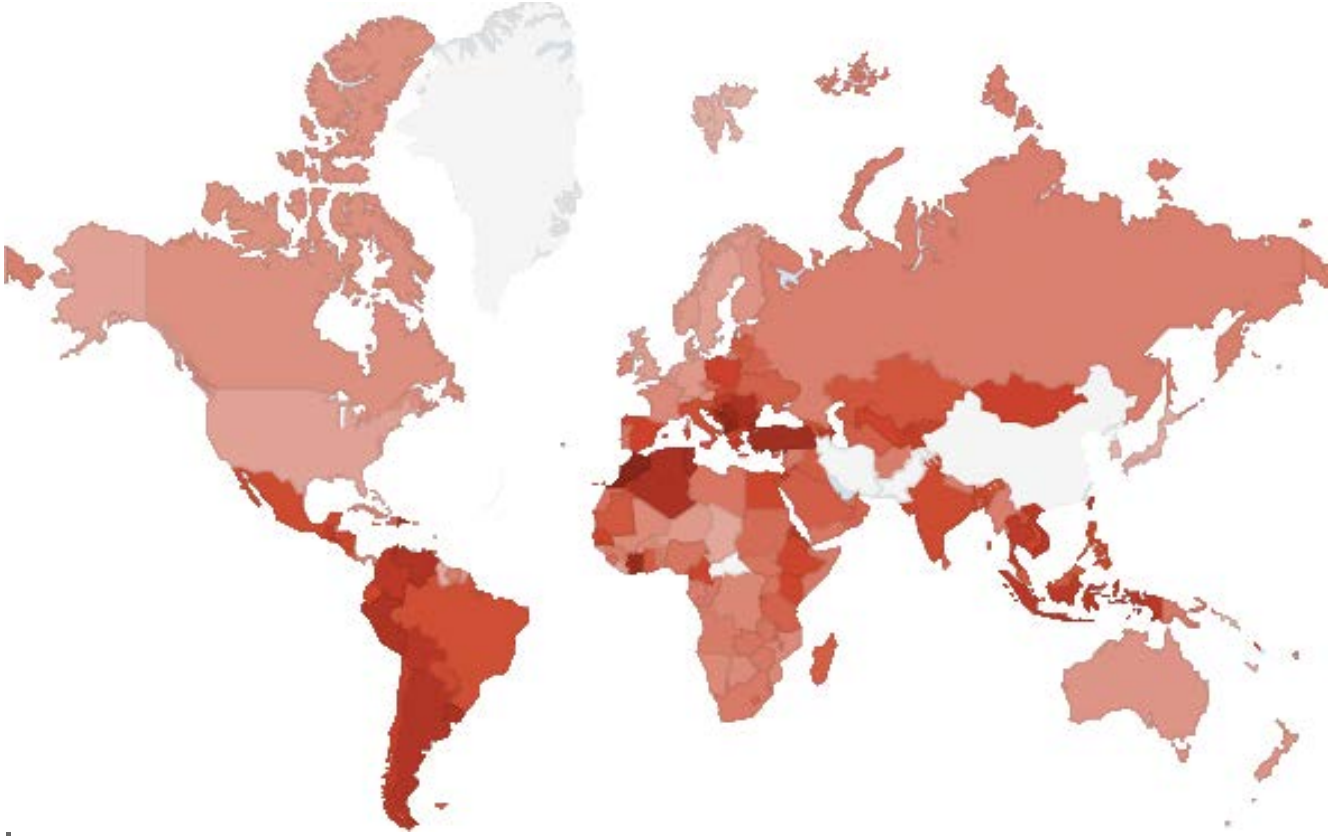
VP9 has improved buffering by 25% in developed markets & 100% in emerging markets

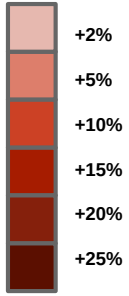
More HD

VP9 has increased YouTube HD delivery by over 25% in emerging markets

Youtube + VP9

VP9





YT VP9 watch time gains vs. H.264

More + VP9

Youtube Live, Google Play Movies & TV, Netflix, …

RTC: AppRTC, Google Hangout, ...

Techinical Preview of AV1

Google

The story continues with AV1

Alliance for Open Media (AOM) is formed in 2015 Q4

AV1

AOMedia Video 1

THE OPEN AND ROYALTY-FREE CODEC FOR NEXT-GENERATION
ULTRA HIGH DEFINITION MEDIA

AOM Formation

Alliance for Open Media

THE OPEN AND ROYALTY-FREE CODEC FOR NEXT-GENERATION ULTRA HIGH DEFINITION MEDIA

Launch: Q4 2015

The Alliance's initial focus is to deliver a next-generation video format that is:

- Interoperable and open;
- Optimized for the Internet;
- Scalable to any modern device at any bandwidth;
- Designed with a low computational footprint and optimized for hardware;
- Capable of consistent, highest-quality, real-time video delivery; and
- Flexible for both commercial and noncommercial content, including user-generated content.

Alliance of Open Media - Members - 2017

The Goal of AV1

Alliance for Open Media

- AV1 Requirements
  - Substantial bitrate reduction for VOD and VC use cases over VP9/HEVC
  - Decoding needs to stay at low enough complexity
    - VOD is the predominant bandwidth consumer for video on the Internet
  - Screen content coding tools
- Achieving a generational improvement in coding efficiency needs:
  - New coding tools, new ways of combining existing/new tools
  - Disclaimer: nothing presented here is set in stone, just a snapshot as of today!

AV1 development

Alliance for Open Media

AV1 started with VP9 tools + enhancements

Development Process:

- Tools from VP10/Daala/Thor/new ones proposed to Codec Workgroup
- Hardware Workgroup vets proposed tools for hardware feasibility
- Testing Workgroup provides a testing framework and conducts psycho-visual tests from time to time
- Review to work towards the goal of royalty-free

Status:

Google

- Many tools adopted - with a decent cumulative gain over VP9

Accessing AV1 Source Code

Alliance for Open Media

- AOMedia git repository:
  git clone https://aomedia.googlesource.com/aom

AV1 Current Status

Alliance for Open Media

- ~40 experimental coding tools, already approved by AOMedia CodecWG
- >20 additional new coding tools are in active development

Planned Milestones for AV1

Alliance for Open Media

- All Major Tools done -> End of October 2017
- Final bitstream -> EOY 2017

AV1 Adoption

Alliance for Open Media

YouTube will begin using AV1 immediately

Netflix will be an “early adopter"
(https://youtu.be/thvSyJN1vsA?t=5452)

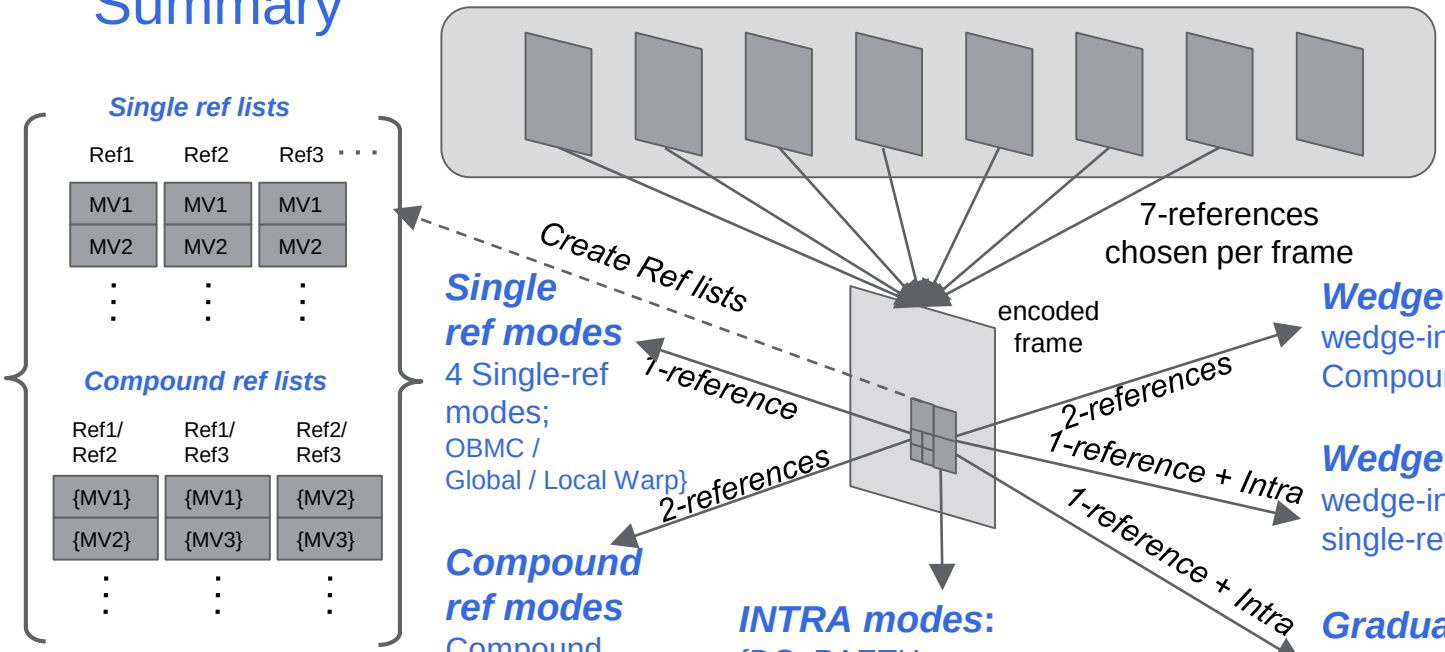
AV1 ecosystem (testing tools + SW optimizations) gearing up

New Coding Tools in AV1

Alliance for Open Media

- Prediction tools
- Transform tools
- In-loop filtering tools
- Miscellaneous tools

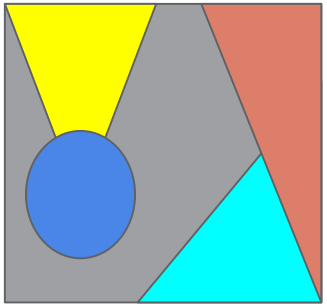
AV1 Technical Overview: Prediction Tools: Summary

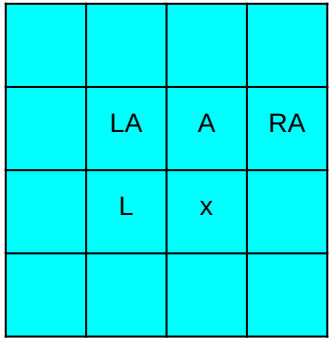


AV1 Technical Overview: Prediction Tools:

INTRA Prediction - Palette mode

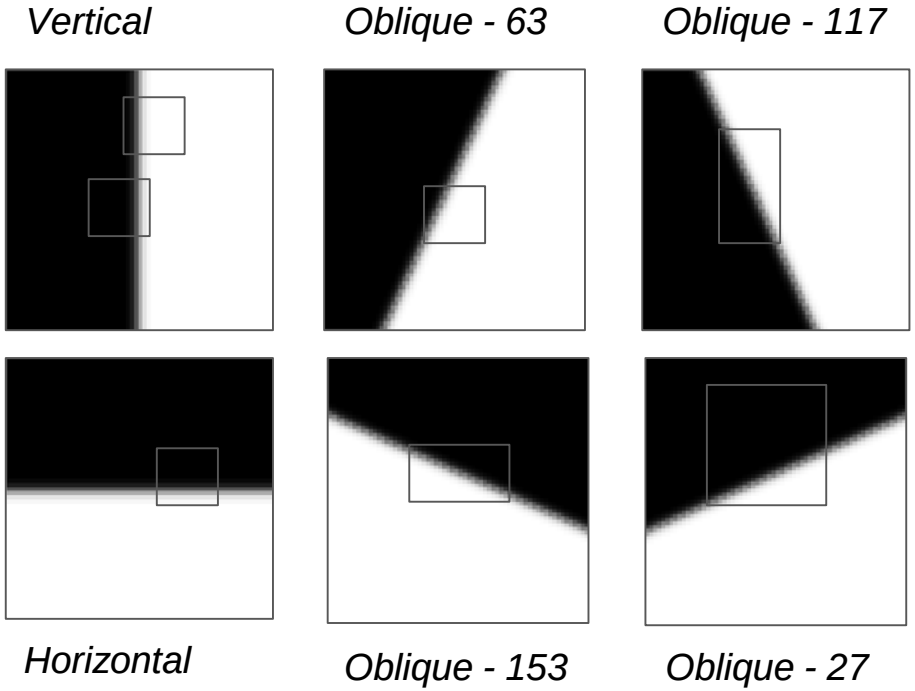
- Particularly useful for screen content
- Palette representation:
  - Encode k distinct *base* colors: k ∈ {2, 3, …, 8},
  - Encode a k-ary color index map
- Base colors:
  - Delta encoded w/ colors from causal neighbor blocks
- Color index map:
  - Use causal neighbor pixels within block as context
- Regarded as a new INTRA prediction mode





AV1 Technical Overview: Prediction Tools:
INTER Prediction: Interinter Wedge Mode

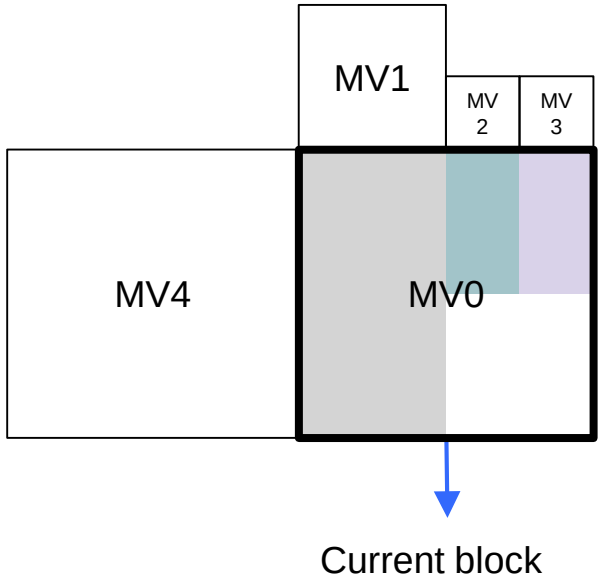
- Wedge mask codebook
  - Send index from codebook
    - Fully determines the mask
- Codebook generation
  - Conveniently generated for each block-size from three length-64 arrays row-shifting
    - On-the-fly generation
- *Overlapped wedge motion compensation*



AV1 Technical Overview: Prediction Tools:
INTER Prediction: Overlapped Block Motion Compensation

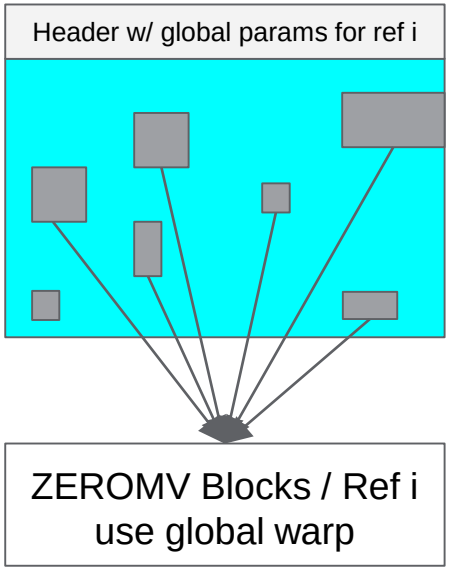
- OBMC
  - Blend multiple predictors from neighboring blocks
  - Not a new concept
    - But H.263 in the 90s was the last codec to use it when block partitioning was much more straightforward

- AV1 reintroduces OBMC
  - Much harder to incorporate in a modern codec and still get gains



AV1 Technical Overview: Prediction Tools:
INTER Prediction: Affine Motion: Global Motion Mode

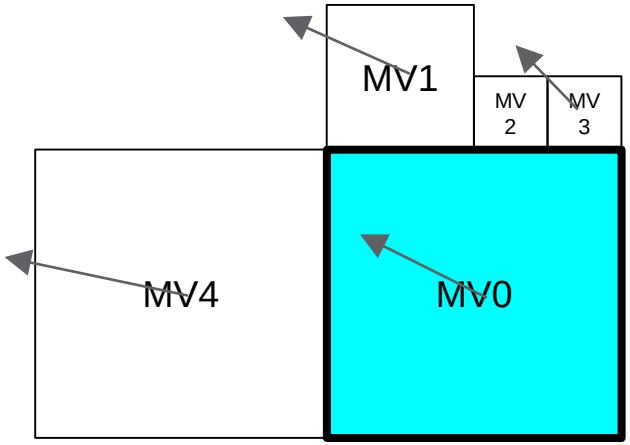
- For each frame in its header for each reference, optionally send:
  - 6-parameter affine parameters
  - 4-parameter similarity parameters (rotation-zoom)
  - 2-parameter translational parameters
- Within the frame if ZEROMV is signaled for a prediction block with a reference that has global parameters, warped prediction is used
  - For sub-8x8 blocks translational motion is used where the block's MVs are computed using global parameters at its center



AV1 Technical Overview: Prediction Tools:

INTER Prediction: Affine Motion: Local Warped Motion Mode

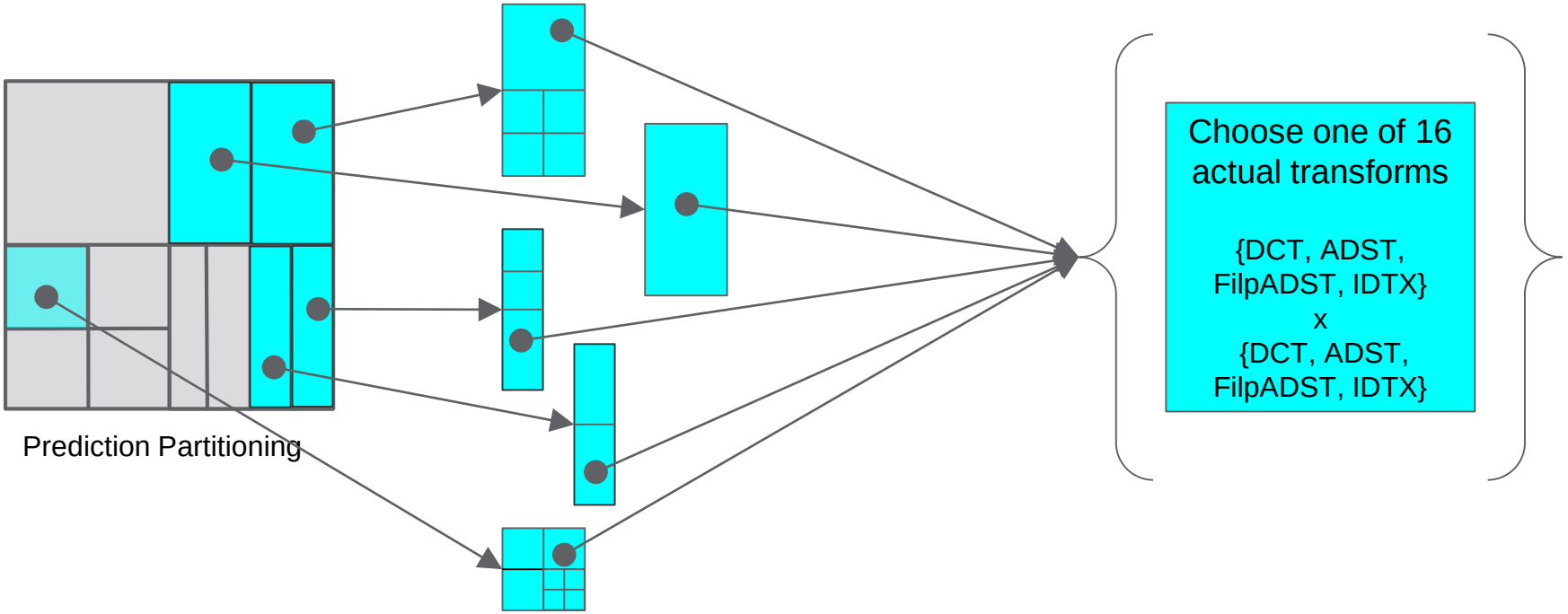
- Goal: Estimate affine parameters using neighboring blocks' MV with same reference.
- Affine least squares:
  - Can be always decomposed into two 3-dim least squares problems
  - MV transmitted is constrained to be consistent with model at the block center
  - This constraint converts each 3-dim problem into a simpler 2-dim problem that is readily integerized.



Blocks 1, 3, 4 have same reference as current block

AV1 Technical Overview: Transform Tools:

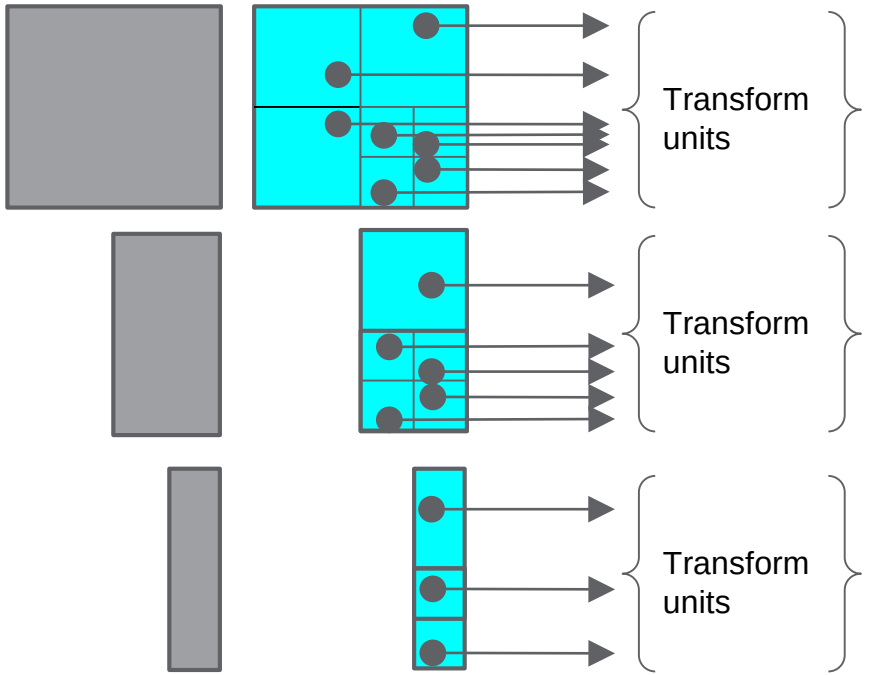
Transform Tools Summary



AV1 Technical Overview: Transform Tools:

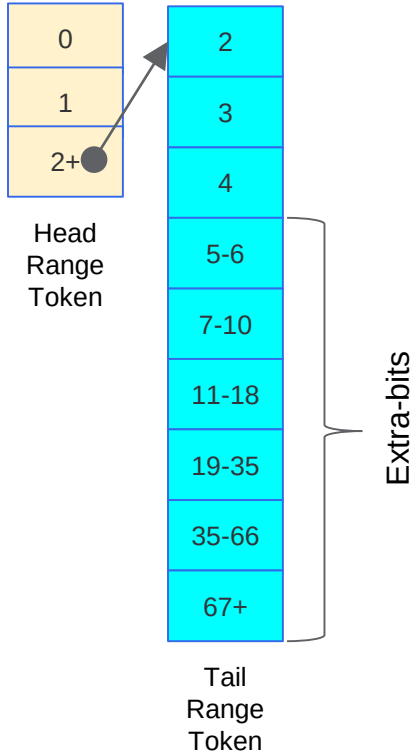
Recursive partitioning of Transform units

- AV1:
  - Allows transforms within a single prediction unit to have quadtree partitions up to two levels
    - More flexibility - can isolate part of a larger block with residue concentration
  - Max transform size still smaller than Partition Unit size
  - Each transform unit has its own transform type



AV1 Technical Overview: Transform Tools:
Coefficient Coding

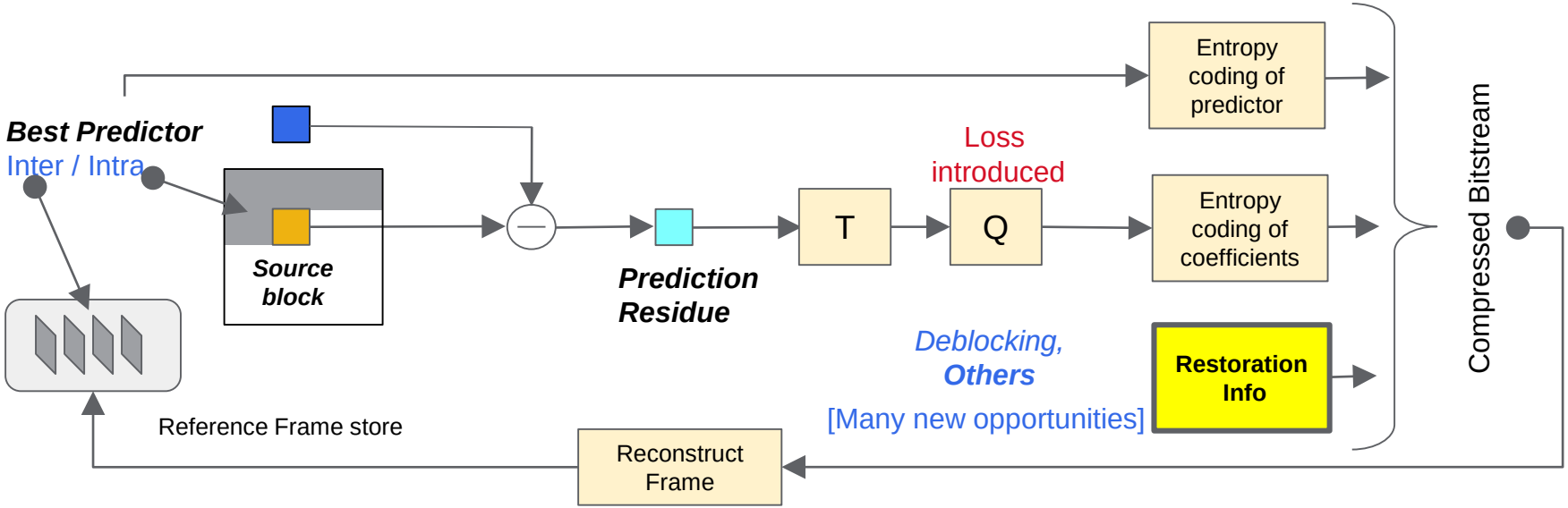
- AV1 Coefficient coding:
  - A coefficient consists of:
    - A Sign
    - Magnitude Range combined with EOB and Zero Block. Has Head and a Tail Token:
      - A Head token from:
        - {BLOCK_ZERO, ZERO, ONE_EOB, ONE_NEOB, TWOPLUS_EOB, TOWPLUS_NEOB}
      - A Tail token: codes other ranges as a 9-valued symbol
    - Extra bits indicating the value within a range
      - Coded 4 at a time in 16-valued symbols



AV1 Technical Overview: In-loop filtering tools

- Many sophisticated image restoration techniques known
  - Deblocking, Deringing, Deblurring, Denoising, Superresolution
- Seldom used for compression
  - Except Deblocking - which is standard
  - Computation is a major concern for video
- Restorability can be equated with compressibility
  - Information that can be restored, can be saved !

AV1 Technical Overview: In-loop filtering tools



- Several tools in AV1 for in-loop restoration other than Deblocking
  - CDEF (CLPF, Deringing)
  - Loop Restoration (Wiener, Dual Self-guided filter)

AV1 Technical Overview: In-loop filtering tools

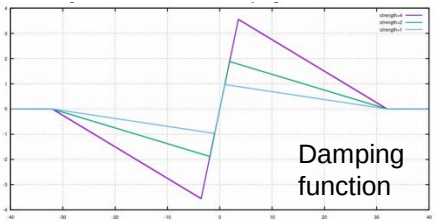
Deblocking filter Advances:

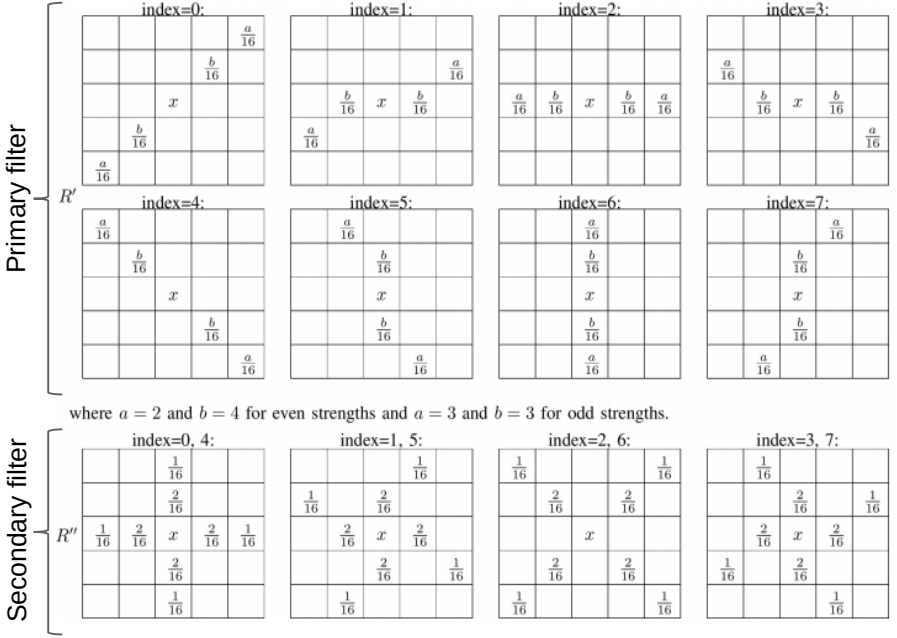
- VP9:
  - One filter level sent per frame
- AV1:
  - Four filter levels per frame:
    - Luma plane: horizontal and vertical filter levels
    - Chroma plane: U, V each has its own filter level.
  - Adds ability to change filter level from superblock to superblock
  - Parallel deblocking [Intel/Microsoft]
    - Apply deblocking for all vertical edges first, then horizontal
    - Adds ability for parallelism

AV1 Technical Overview: In-loop filtering tools

CDEF [Mozilla/Cisco]

- Direction-adaptive nonlinear filtering in CDEF:
  - Filter expressed as difference to center pixel of 5x5 support region
    - Differences go through nonlinear damping function f() before being filtered through a *primary* and a *secondary* 2D filter
  - Sum the two filter outputs, clip to the max difference within the 5x5 block, add to the center pixel.

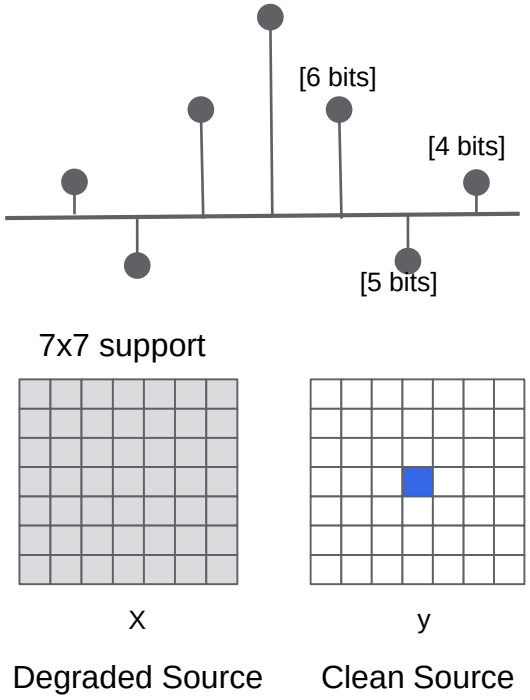




AV1 Technical Overview: In-loop filtering tools

Loop-Restoration: Wiener Filter

- Separable Symmetric Normalized Wiener filter:
  - 7-tap horizontal and 7-tap vertical separable
  - Symmetric, Normalized
  - ~30 bits per RU, Easy for decoder
- Encoder:
  - Design filters s.t. when applied to a deblocked frame produces an output closer to source
  - Design:
    - Classical non-separable LMMSE design: $H = R_{xx}^{-1} R_{xy}$ not possible
    - Impose separability, symmetry and normalization constraints
    - Iterative design for hor/vert filters

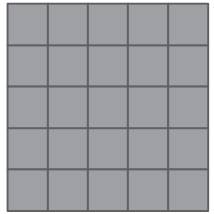


AV1 Technical Overview: In-loop filtering tools

Loop-Restoration: Dual Self-guided filter

- Guided image filtering [2010]
  - Fit a local linear model considering a guide image
    - Self-guided: Guide image = Input Image
  - Can be used as an edge-preserving smoother
- Steps with parameters (r, e) in AV1:
  - Obtain mean μ and variance σ^2 of pixels in a (2r + 1) x (2r + 1) window around every pixel.
  - Compute: $f = \sigma^2/(\sigma^2 + e)$; $g = (1 - f)\mu$
  - Compute F, G by averaging f, g in a 3x3 area around each pixel.
  - Filter pixel value using: y = x. F + G
  - Efficient implementation possible: readily integerized / vectorized

Parameters: (*r*, *e*)
r can be 1 or 2



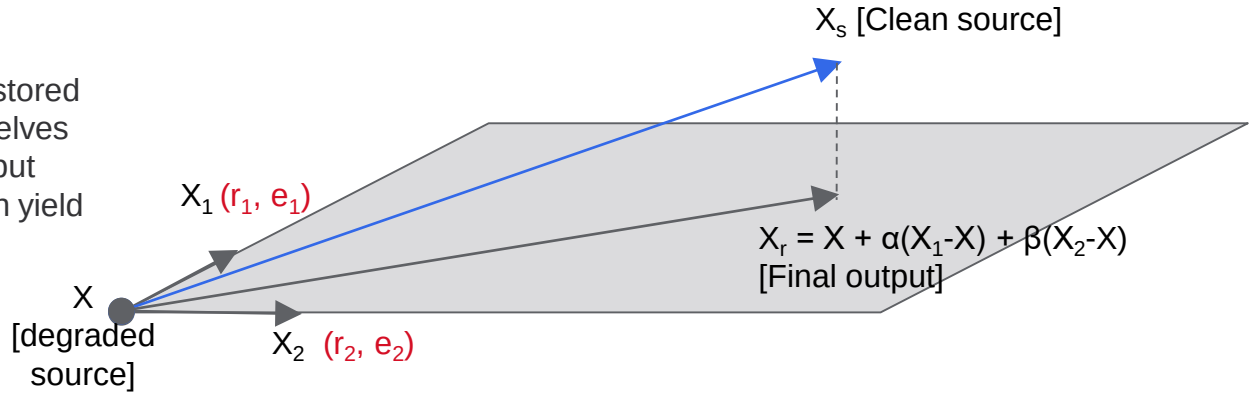
(2*r* + 1) x (2*r* + 1)

AV1 Technical Overview: In-loop filtering tools

Loop-Restoration: Dual Self-guided filter

- Dual Self-guided filter with subspace projection in AV1:
  - Filter twice: Once with (r_1, e_1) yielding X_1, and once again with (r_2, e_2) yielding X_2
  - Use weighted combination w/ weights {α, β} of X_1, X_2, to get final output
  - (r_1, e_1, r_2, e_2) taken from a small 4-bit codebook ($r_1 = 1$; $r_2 = 2$);
  - {α, β} uses 7-bits each
  - ~18 total bits per RU

X_1 and X_2 are cheap restored versions, and by themselves may not be very good; but subspace projection can yield a much better final restoration X_r.



AV1 Technical Overview: Miscellaneous tools

Tiling Features

- Motivation:
  - Parallel encoding/decoding for speed
  - Random Access: Allow decoder to extract only an interesting section in a frame
  - Error-resilience
- AV1 tiling features:
  - Independent tiles so that tiles in a frame can be encoded / decoded in parallel.
  - Tile width ranges from 64 pixels to 4096 pixels.
  - Tile area limited to a maximum of 4096 x 2304 pixels.
  - Support flexible tile, including uniform and non-uniform tile spacing.
  - Tiles can be grouped into *tile_groups* and each *tile_group* is independently decode-able to achieve error resilience.
  - Loop-filtering can be enabled or disabled across tiles.

AV1 Technical Overview: Miscellaneous tools

Hardware Simplification Contributions

- Several contributions from various partners to simplify hardware
  - MV referencing simplifications
  - 7-bit filter taps
  - Deblocking + CDEF + LR pipeline

Preliminary Results

AWCY

- www.arewecompressedyet.com
  - An automatic service for compressing videos and comparing coding efficiency among AV1, VP9, Daala, Thor, and other codecs.
    - Maintained by: xiph.org foundation [Mozilla]
  - Chosen as official testing infrastructure for AOM
  - Uses AWS Cloud resources
  - Supports several test sets, and test conditions

Preliminary Results

AWCY: High Latency

- Objective-1 Fast - High Latency
  - Constant Quality mode, 30 clips w/ 60 frames
- Candidates
  - Baseline: VP9 tip of tree; Test: AV1 (08/26/2017), X.265 ver 1.9

BDRATE	PSNR-Y	PSNR-Cb	PSNR-Cr	PSNR-HVS	CIEDE 2000	SSIM	MS-SSIM
AV1*	-25.58%*	-21.78%	-23.69%	-24.14%	-24.43%	-23.63%	-22.58%
X.265+	-4.46%	N/A	N/A	-2.25%	+13.33%	+8.24%	+7.88%

+X.265: --preset placebo --tune psnr --no-wpp

*AV1 target: -30% over VP9

Preliminary Results

AWCY: Low Latency

- Objective-1 Fast - Low Latency
  - Constant Quality mode, 30 clips w/ 60 frames
- Candidates
  - Baseline: VP9 tip of tree; Test: AV1 (09/15/2017)

BDRATE	PSNR-Y	PSNR-Cb	PSNR-Cr	PSNR-HVS	CIEDE 2000	SSIM	MS-SSIM
AV1*	-29.31%*	-28.43%	N/A	-27.11%	-28.10%	-30.48%	-28.44%

*AV1 target: -30% over VP9

Conclusion

- AV1 Coming Soon!
  - AOM has made decent progress towards a next generation royalty-free codec **AV1**
    - 25+% BDRATE reduction over VP9 and HEVC at high delay
    - 29+% BDRATE reduction over VP9 at low delay
    - With high-bit-depth internal there is an additional 3-4% gain
    - Our target: 30%
  - Google VP10 effort evolved into AV1
  - Expected AV1 freeze: December 2017
  - Encoder / decoder optimization will continue

Thank you!