Progress in the Alliance for Open Media

Timothy B. Terriberry

Mozilla & The Xiph.Org Foundation

Setting the Stage

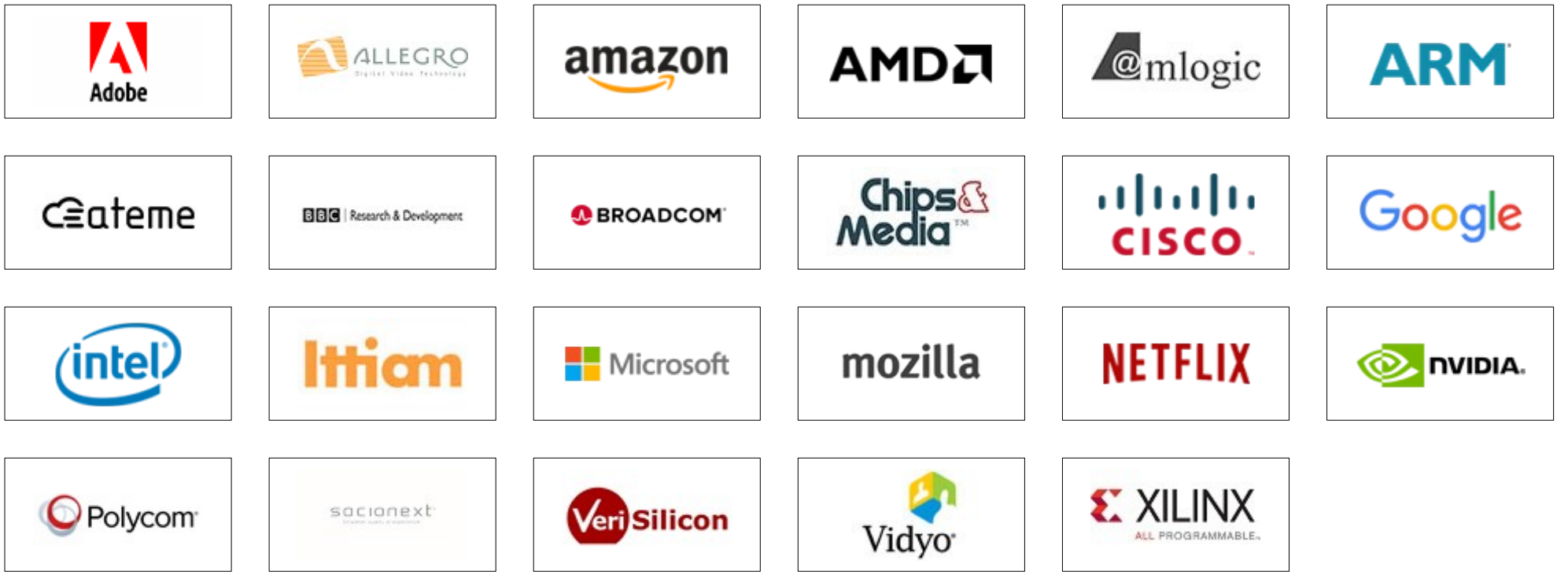
- Early 2015
  - Google hard at work on VP10, successor to VP9
  - Spoke in Auckland on the successes we (Mozilla/Xiph) were having with Daala
  - Cisco open-sourced Thor, a simple royalty-free codec aimed at real-time
- Someone at my talk rightly asked:
  - “So, are you guys talking to Google?”

Something Happened

- MPEG LA announces annual caps for HEVC almost 8× higher than H.264
- HEVC Advance decides that isn't enough
  - New pool with per-unit fees up to **10× higher**
  - **No annual cap**
  - Also adds content fees (0.5% of revenue)
- Annual costs would be 100× those of H.264
  - For someone of Mozilla's size
- That got people's attention

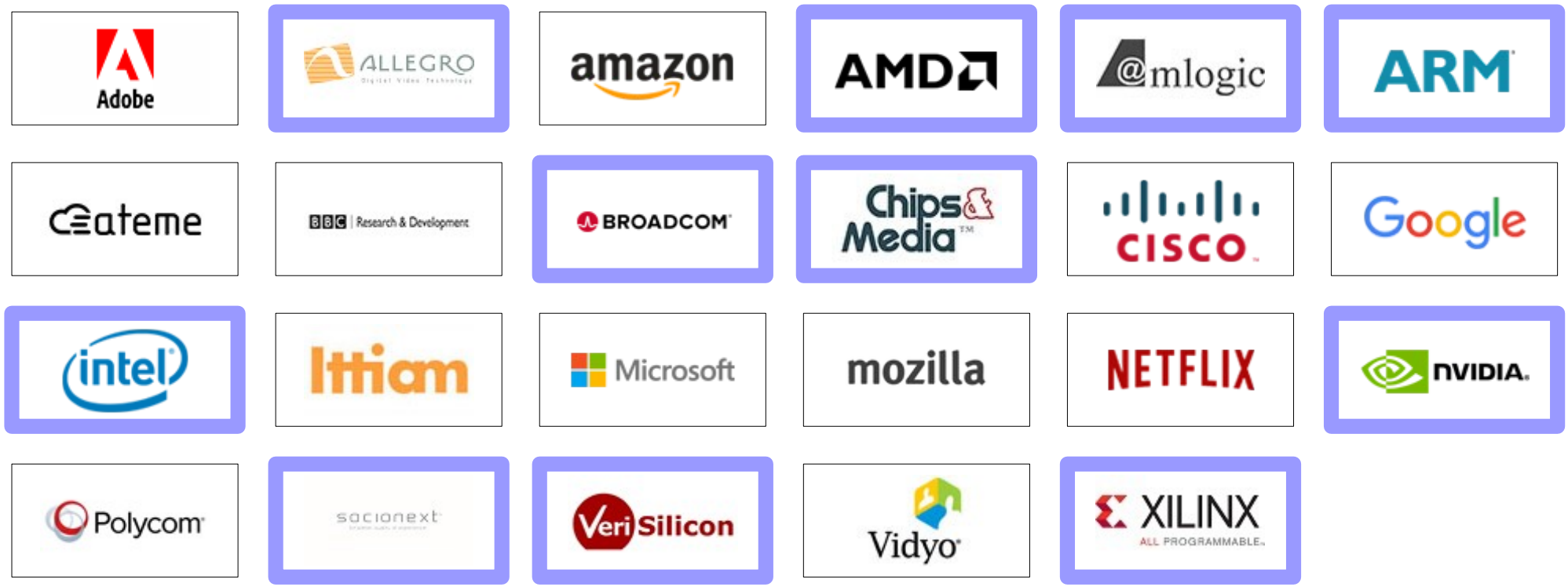
Enter the Alliance for Open Media

- Formed by 8 companies in September 2015
- Now we have more



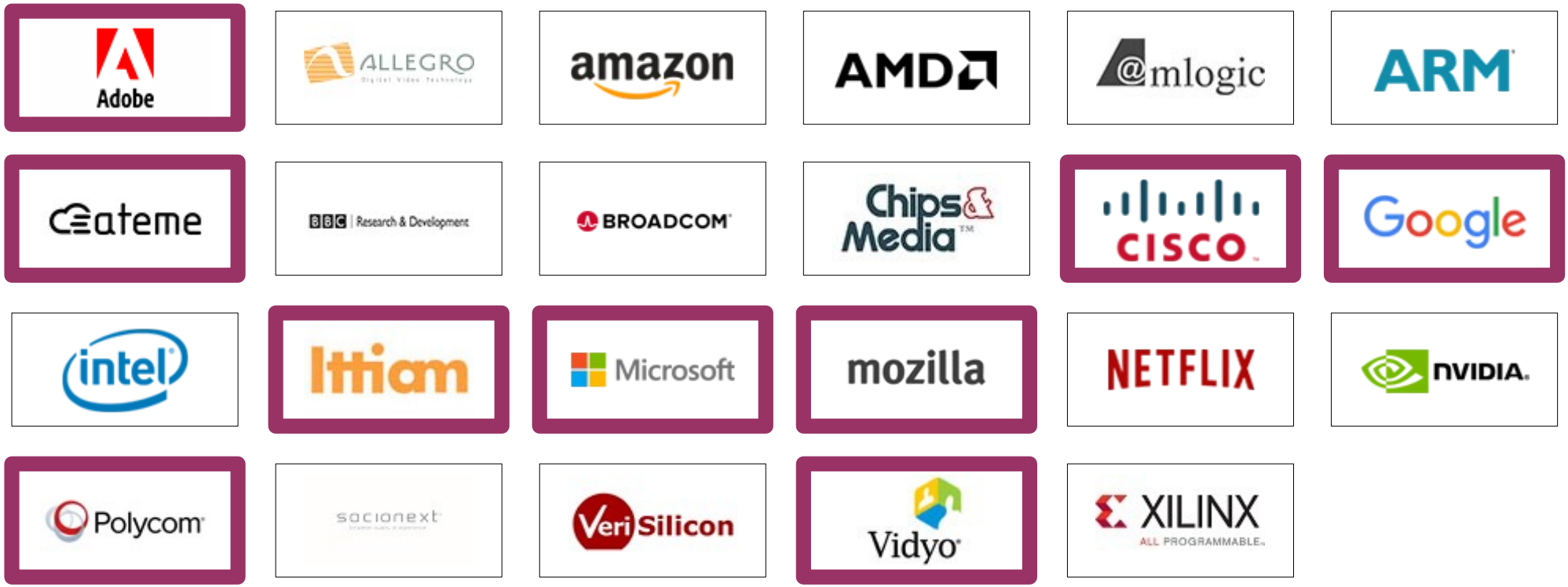
The Alliance for Open Media

- Some of us make hardware



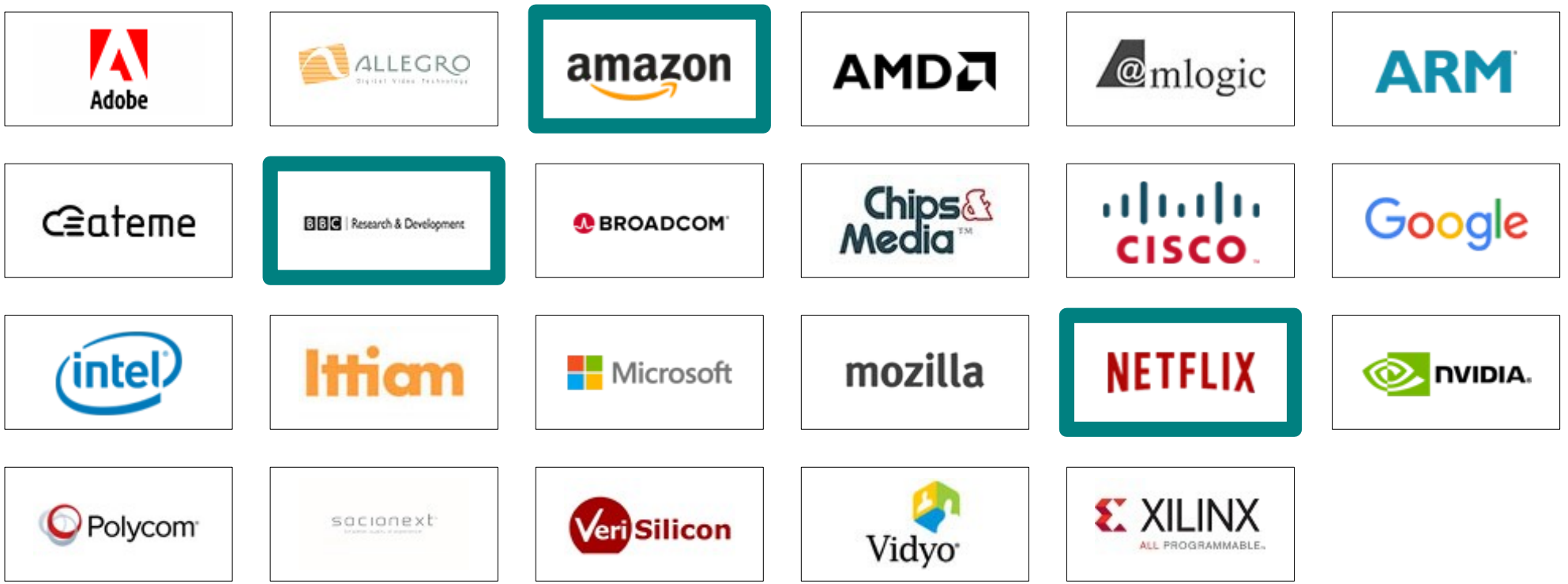
The Alliance for Open Media

- Some of us make software



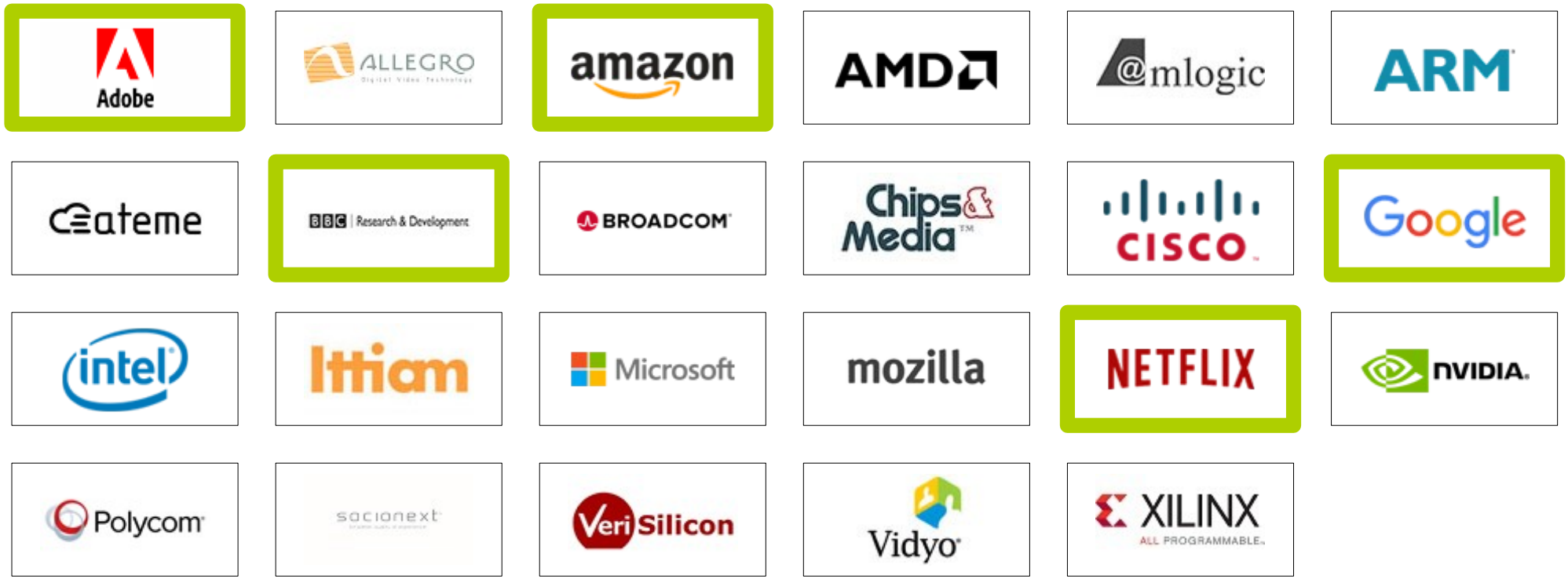
The Alliance for Open Media

- Some of us produce content



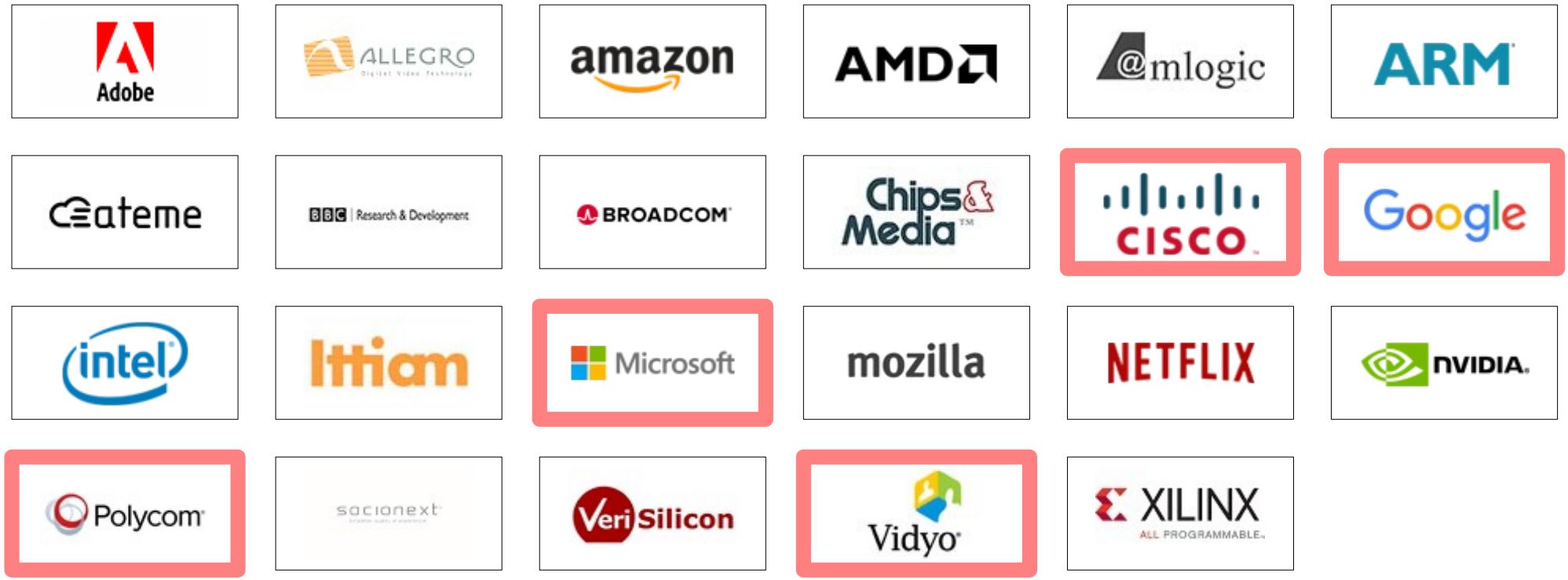
The Alliance for Open Media

- Some of us stream content



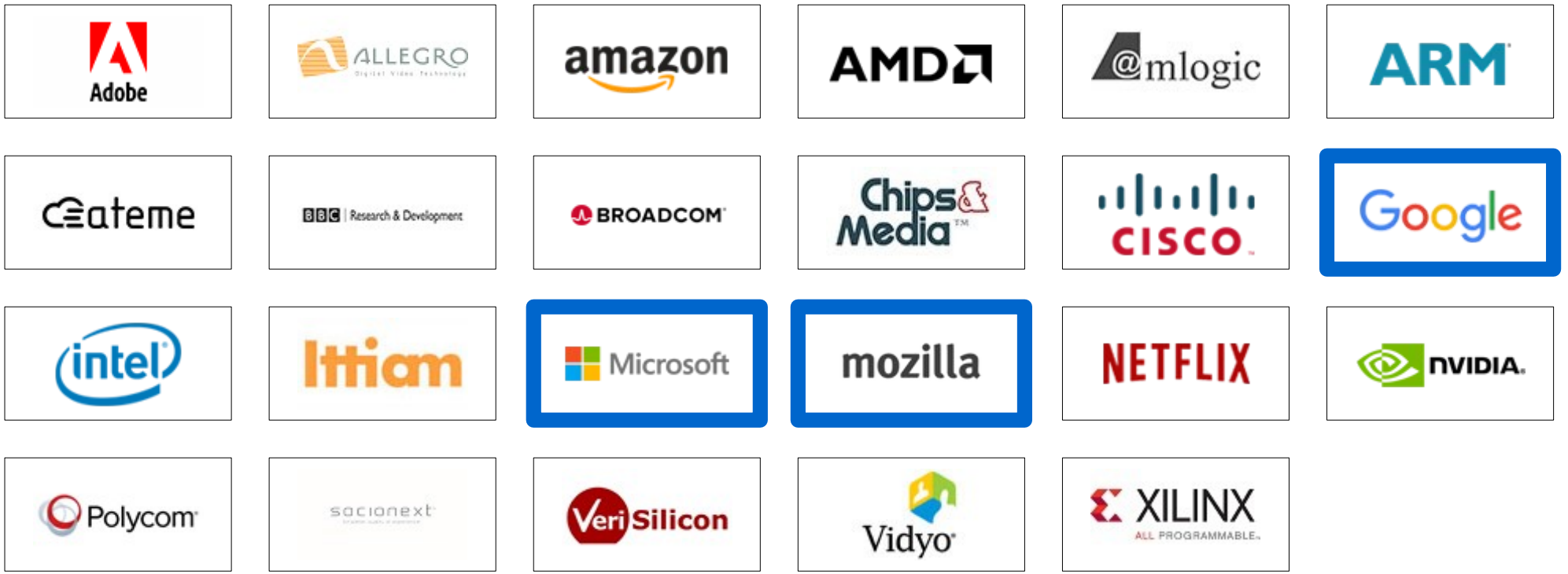
The Alliance for Open Media

- Some of us do real-time conferencing



The Alliance for Open Media

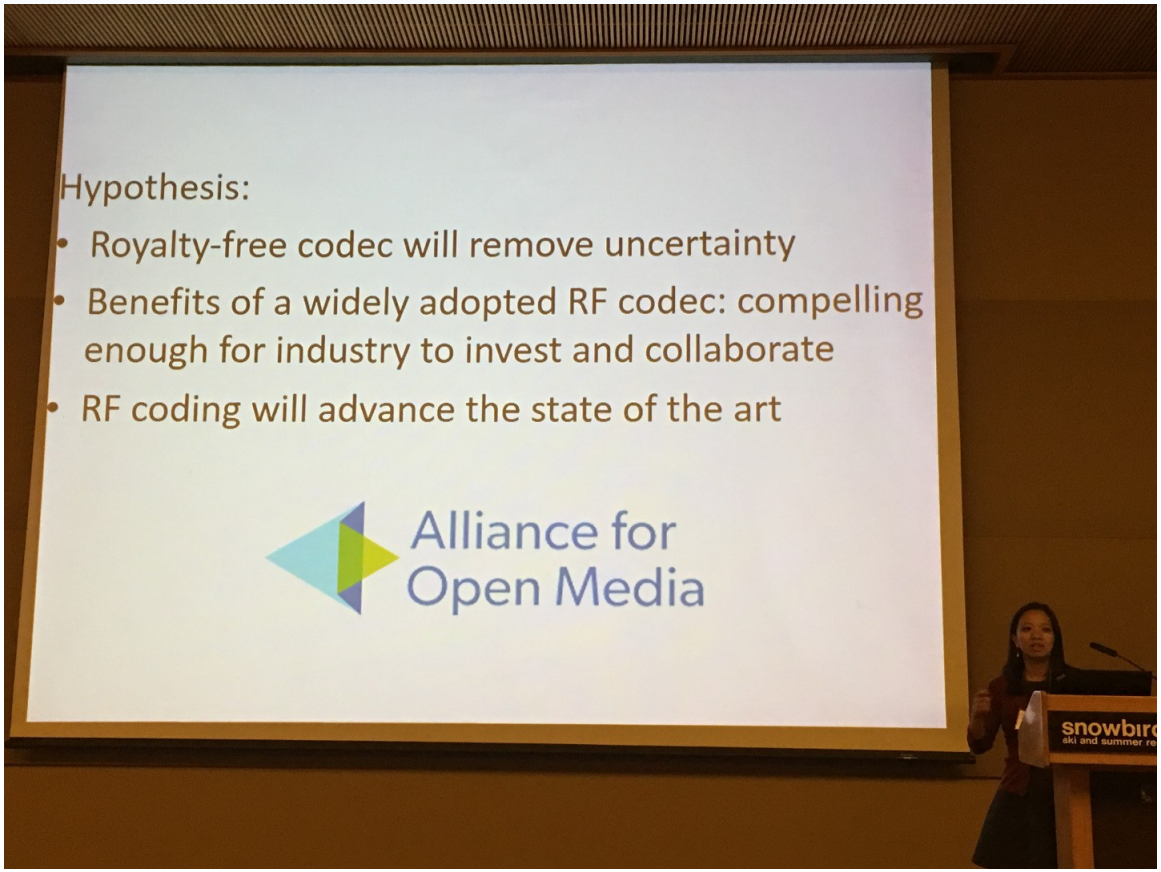
- Some of us make browsers



MPEG Model Not Working

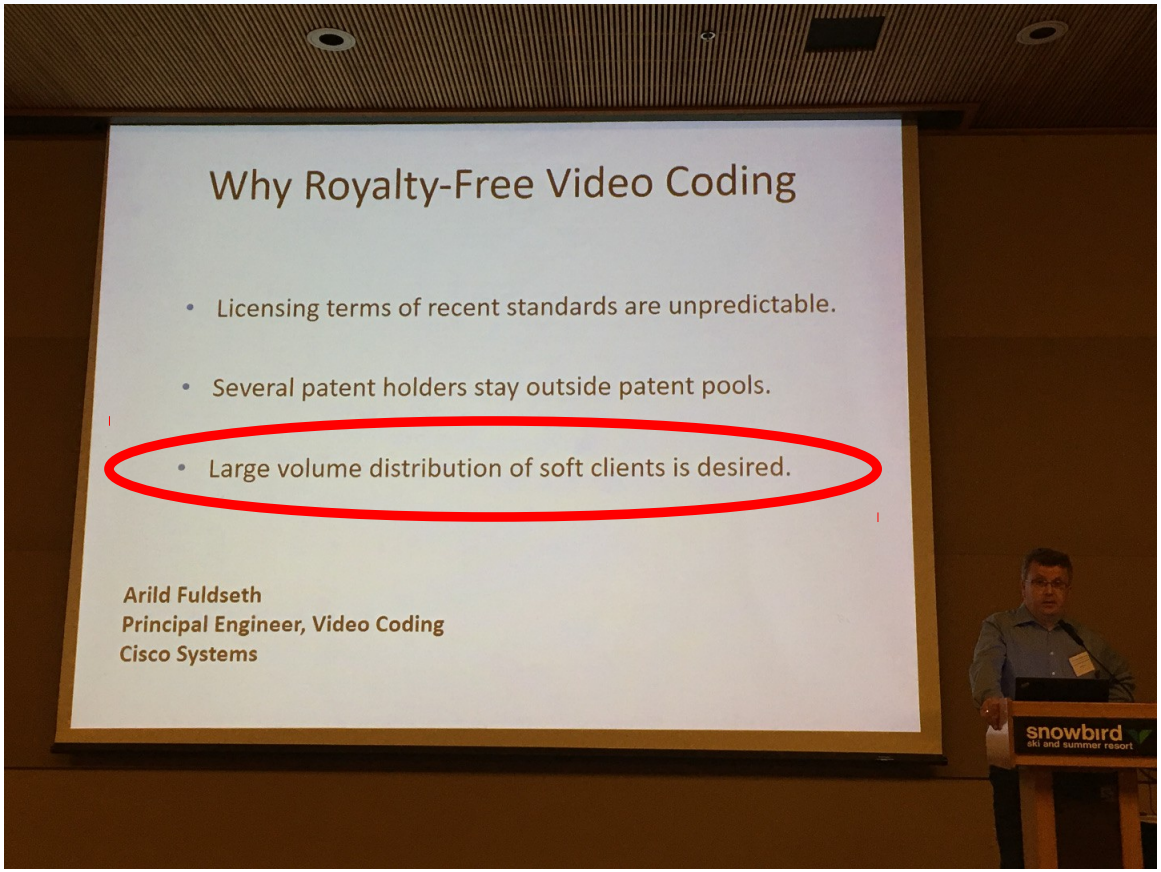
- The idea of “develop the technology first, figure out the licensing later” is broken
  - Guarantees a patent-encumbered mess
  - Subject to gaming
    - And people have figured out how over 30 years
  - Vulnerable to “patent hold-up”
    - Individuals want more than the technology is worth, because they think you’ll have to pay
    - Many companies can’t afford to use the technology they helped create
- And it’s not just us saying so anymore

Not Just Us...



Anne Aaron of Netflix at Data Compression Conference 2016

Not Just Us...



Arild Fuldseth of Cisco at Data Compression Conference 2016

So, What's AOM's Patent License?

- Negotiated a single royalty-free license granted by all AOM members
  - Better than Opus, with a Xiph/Broadcom license and a separate Microsoft license
  - Combines elements of both
- Covers both decoder *and encoder*
  - Anything required by the specification or included in the reference implementation
- http://www.aomedia.org/license/patent

So, What's AOM's Patent License?

- Includes a defensive termination clause
  - If you sue *anyone* over using our codec, you lose our license
- Includes a reciprocal license grant
  - Uses BSD mechanism as the affirmative act that signals license acceptance
    - Include a copy with source code/product documentation
  - Open-Source friendly (this is novel)

Isn’t This a Solved Problem?

- We have VP9, and it does pretty well, but..
  - 4k is 4× larger than 1080p
    - 8k is 4× larger than that
  - 60 fps is 2× larger than 30 fps
  - High Dynamic Range is 25-50% larger than SDR
- That’s a lot more data, and a lot more bits
- And video is 70% of all consumer internet traffic
  - So compressing it better is a big deal

Science


MISSION DAY
SOL 71
PRESSURE
12.48
PSI
OXYGEN
20.78
%
TEMP
21.06
C
TIME 19:33
LOG ENTRY > WATNEY #113

How AV1 Development Works

- Fully open-source
  - https://aomedia.googlesource.com/aom/
  - Don’t have to join AOM to contribute
- Specific coding tools added as “experiments”
  - Controlled at build-time by flags
  - `--enable-experimental --enable-<foo>`
- Experiments get review by hardware team, IPR team (TAPAS), and whole working group
  - Default disabled → Default enabled → Flags removed

Experiments

- About 50 active experiments
- 5 currently enabled by default
- Will mostly focus on Mozilla's contributions
  - Daala's strategy: look for common elements in lots of patents, replace with something different
  - Room to get some of these replacements into AV1
  - Doesn't have to be as complete as Daala
    - AOM members have patents, too
    - Lots of members: more experience, shared IPR review
    - AOM has money to pay lawyers

Binary Entropy Coding

- Most codecs only code binary decisions
  - Actual cost in bits depends on probability
  - Very cheap to code 1 symbol
- Binary probability modeling
  - Simple 1-byte lookup tables
- Various schemes for converting non-binary values to binary decisions (“binarization”)
  - Need to code a lot of symbols (not parallelizable)

Non-Binary Entropy Coding (`ec_multisymbol`)

- Code values with up to *16* possibilities
  - Equivalent to 4 binary decisions
  - More expensive, but not 4× more expensive
    - A lot of overheads are *per-symbol*
  - Effectively parallel!
- One byte cannot model 16 probabilities
  - Modeling more expensive, but easily SIMDable
- Convert things to hex, not binary!
  - Often combine multiple values into one symbol

Non-Binary Entropy Coding (`ec_multisymbol`)

- Fewer symbols means
  - Shorter serial dependency chain
  - Hardware can use lower clock rates
  - Devices use less power
- Two new underlying engines
  - `daala_ec`: the current default
  - `ans`: faster in software, complications for hardware/realtime (must encode in reverse order)
- Probabilities currently converted from binary probabilities

Adaptive Entropy Coding (ec_adapt)

- Work lead by Cisco (based on Daala approach)
- VP9 has two ways to control probabilities
  - Explicitly send probabilities in the frame header
    - Takes a lot of bits, especially at low resolutions
    - Need to encode whole frame to pick best probabilities
      - Adds latency
      - Must make decisions without final probabilities (suboptimal)
  - Update probabilities from previous frame statistics
    - Not robust to packet losses
    - Limits frame parallelization

Adaptive Entropy Coding (ec_adapt)

- Adapt probabilities per symbol:
  - Already doing most of the work
    - Updating once per frame required symbol statistics
  - Cost scales with symbols coded
    - Per-context learning overhead approximately log(*N*) bits
    - Lower rates/lower resolutions → less overhead
  - Can adapt to non-stationary statistics

	PSNR	CIEDE 2000	PSNR HVS	SSIM	MS SSIM
High Delay	-0.38%	-0.68%	-0.72%	-0.67%	-0.78%
Error Resilient	-1.66%	-1.55%	-2.60%	-2.12%	-2.47%

- Not updating during RDO: can do better

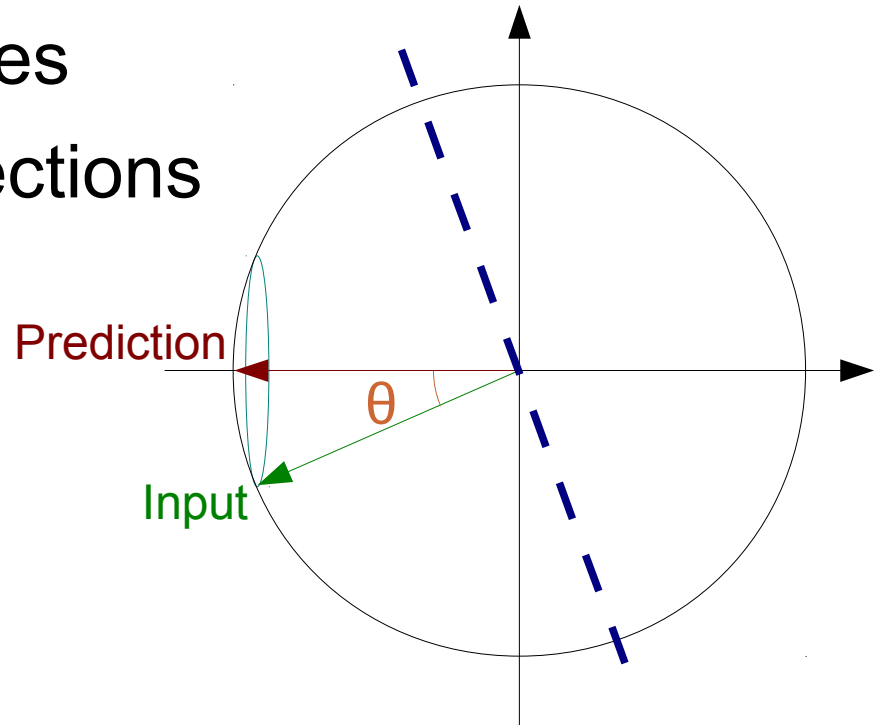
Adaptive Deringing (dering)

- Daala has a lot of ringing artifacts
- We made a pretty good filter to reduce them
  - Estimates dominant orientation in each block
  - Strong smoothing parallel to the edge
    - And then weaker smoothing orthogonal to it
  - Signal strength on 64×64 block basis (can shut off)
- Works fine in AV1

	PSNR	CIEDE 2000	PSNR HVS	SSIM	MS SSIM
High Delay	-1.52%	-1.26%	-0.89%	-1.47%	-1.01%
Low Delay	-3.01%	-2.61%	-2.41%	-2.05%	-2.21%

Perceptual Vector Quantization (pvq)

- Replaces computing displaced frame difference and scalar quantization
- Not going to explain all the details
  - It involves hyperspheres
  - And Householder reflections
- Will it work outside Daala?



PVQ: Initial Integration

- Completed in November
- Optimizing for PSNR
  - Just for comparison purposes
  - Have much more “perceptual” metrics to target
- Overall results:
  - 0.1% worse PSNR (basically noise)
  - 0.94% better to 1.27% worse across all metrics

PVQ: Challenges

- Recent regressions
  - Now 3...4% worse, still tracking down why
- Speed: 20× slower encode
- Why?
  - Missing SIMD (maybe 3...4× improvement)
  - Requires extra forward transform
  - Slower search than scalar
  - But all that was true in Daala, and it wasn't 20× slower...

Quantization gets called a *lot*

- The AV1 search space is huge
- The encoder makes decisions by transforming, quantizing, and encoding for almost everything

Cost of coding a single 176×144 frame

AV1 Speed Level	Number of calls to PVQ
5	26,786
4	56,980
3	564,724
2	564,724
1	580,566
0	632,520
Daala	3,843

← 165×

This is going to get worse

- Some other experiments in the works...
  - `ext_tx`: Number of transforms: 4 → 16 [**4×**]
  - `ext_intra`: Number of intra prediction modes: 10 → 58 [**5.8×**]
  - `ext_partition_types`: 4 splitting modes → 8 [**2×**]
  - `ext_refs`: 3 references per frame → 6 [**2×**]
  - `ext_inter`: mixed inter/intra in a single block, “wedge” predictors, etc.
  - About **65×** slower to encode with all this... without PVQ

So what can we do?

- Stop calling the quantizer so much
  - Search less of the space (w/heuristics, etc.)
    - Some of the experiments already do this
  - Replace transform + quantization + coding with a simple model
    - Make most decisions with the model
    - Call quantizer for final encode
    - Maybe also a few alternatives if decisions are close

Rate-Distortion Modeling

- Measure something simple/cheap to compute
  - E.g., MSE, SAD, SATD of prediction
- Add a few extra parameters
  - Quantizer strength, block size, color plane, prediction type
- Use it to estimate rate (bits) and distortion
  - Lookup table/simple math
- Doesn’t have to be accurate
  - Just has to make accurate trade-offs

Rate-Distortion Modeling

- This benefits the whole codec, not just PVQ
  - We'll need something like this to be practical, regardless of whether or not we use PVQ
- Usually done after standardization
  - We care about implementations people can use
  - We care about time to market
  - If a tool only gives gains with an impractically slow search, is it a useful tool?
- Still a WIP, results soon...

Putting the “Perceptual” in PVQ

- Optimizing PSNR is not what PVQ was designed for
- Want to optimize for looking better to humans
  - Non-flat quantization matrices
    - Humans are less sensitive to high frequencies
  - Activity masking
    - Humans are more sensitive to errors in low-contrast regions (higher *relative* error)
- Doing this will hurt PSNR, help perceptual metrics (and look better)

First Step: A Better Distortion Function (`daala_dist`)

- Daala's distortion function attempts to approximate PSNR-HVS
  - The "HVS" stands for "Human Visual System"
- Chop into 8×8 blocks, weight errors by frequency
- Measure contrast, and increase error in low-contrast regions
- Penalize large differences is contrast (energy)
- Slower than MSE/PSNR
  - But that doesn't matter if most decisions are made by modeling

First Step: A Better Distortion Function (`daala_dist`)

- If your quantization is smart (PVQ), but your distortion function is dumb (PSNR), you'll have a bad time
  - The two will fight
  - You'll make bad decisions at a macro scale
    - Wrong block size, wrong prediction type, etc.
- Not yet landed, but already showing good perceptual gains (without PVQ)

PSNR	CIEDE 2000	PSNR HVS	SSIM	MS SSIM
4.86%	5.38%	-0.48%	-3.12%	-4.56%

AV1 Reference (Low Bitrate)

Daala Distortion (Low Bitrate)

AV1 Reference (Low Bitrate)

Daala Distortion (Low Bitrate)

PVQ: Summary

- Recent regressions aside, comparable PSNR
- But that's not why we want to use PVQ
  - We want to make our IPR review job easier
  - We want to improve how video looks
    - Just starting to integrate those pieces
- Have a plan for achieving reasonable complexity
  - But like any plan, there is some risk

Working on lots more...

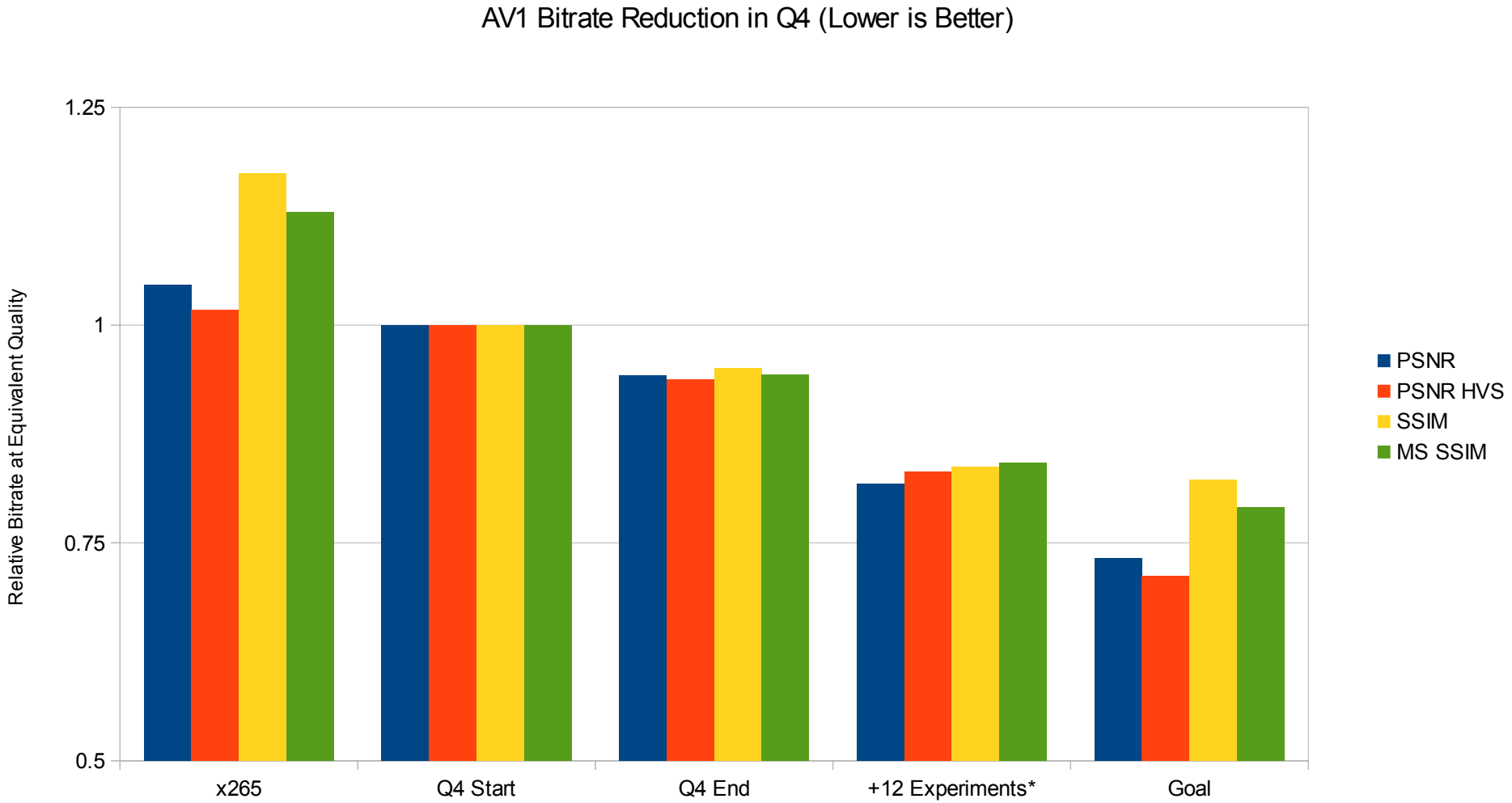
- Improved rate control
  - Supports chunked two-pass with same quality as whole-file encode
- Chroma from Luma
  - Simple extension of PVQ for much cleaner color
- Better packet loss robustness
- Better parallelism
- Emscripten-based bitstream analyzer
  - http://aomanalyzer.org/
- And more...

How Are We Doing?

New HEVC Advance Terms

- A few months after the formation of AOM, HEVC Advance announced revised terms
  - (Somewhat) reduced per-unit fees
  - Added annual caps
- Still 10× more expensive than H.264
  - But for some, that's progress!
- Technicolor left HEVC Advance
  - You must negotiate to license their patents directly
- ~1/3rd of HEVC patent holders not in any pool

Oh, and the codec, too



*ref-mv, ext-intra, var-tx, ext-refs, ext-interp, supertx, motion-var, ext-inter, ext-tx, entropy, loop-restoration, dual-filter, filter-intra

Contributing

- Bug tracker:
  https://bugs.chromium.org/p/aomedia
- Contributor's Guide:
  http://aomedia.org/contributor-guide/
- Contributor Agreement:
  http://www.aomedia.org/license/individualCLA
  http://www.aomedia.org/license/corporateCLA
- Joining AOM: membership@aomedia.org
  - Membership fee required (used for legal analysis)

Questions?