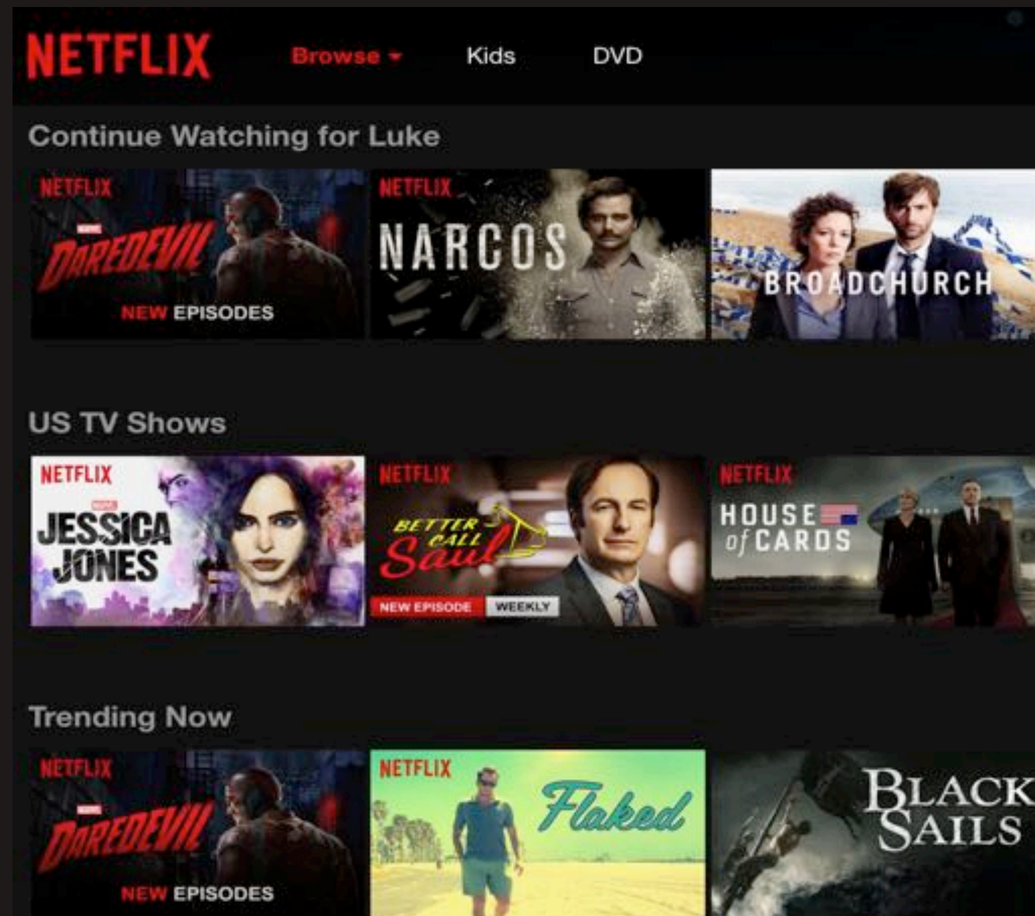
Video codec comparison using the dynamic optimizer framework

Ioannis Katsavounidis &
Liwei Guo
Video Algorithms
Netflix

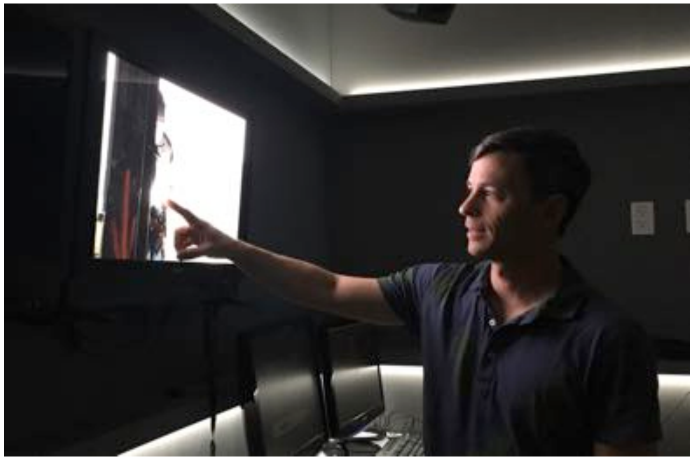
World's Leading Internet Entertainment Service



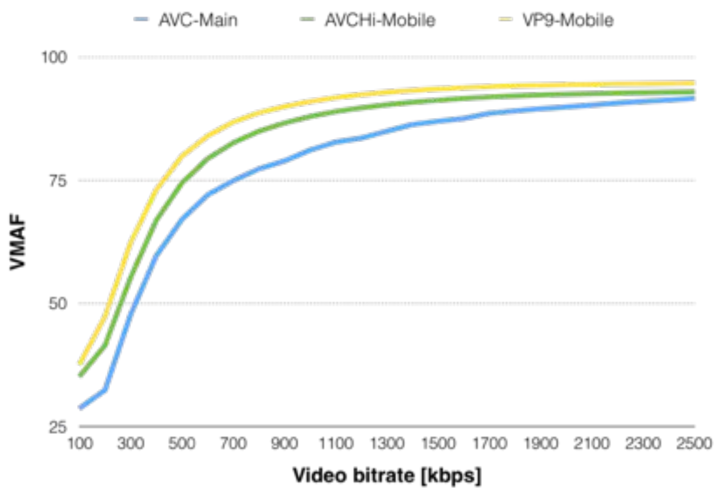
Outline

- Measuring video quality – VMAF
- Dynamic Optimizer framework
- Proposed testing methodology
- Results

Ways to measure video quality



Subjective Assessment



Automated Assessment
using PSNR, SSIM, or VMAF


PSNR 37.3 dB


PSNR 32.9 dB

Need a better perceptual metric

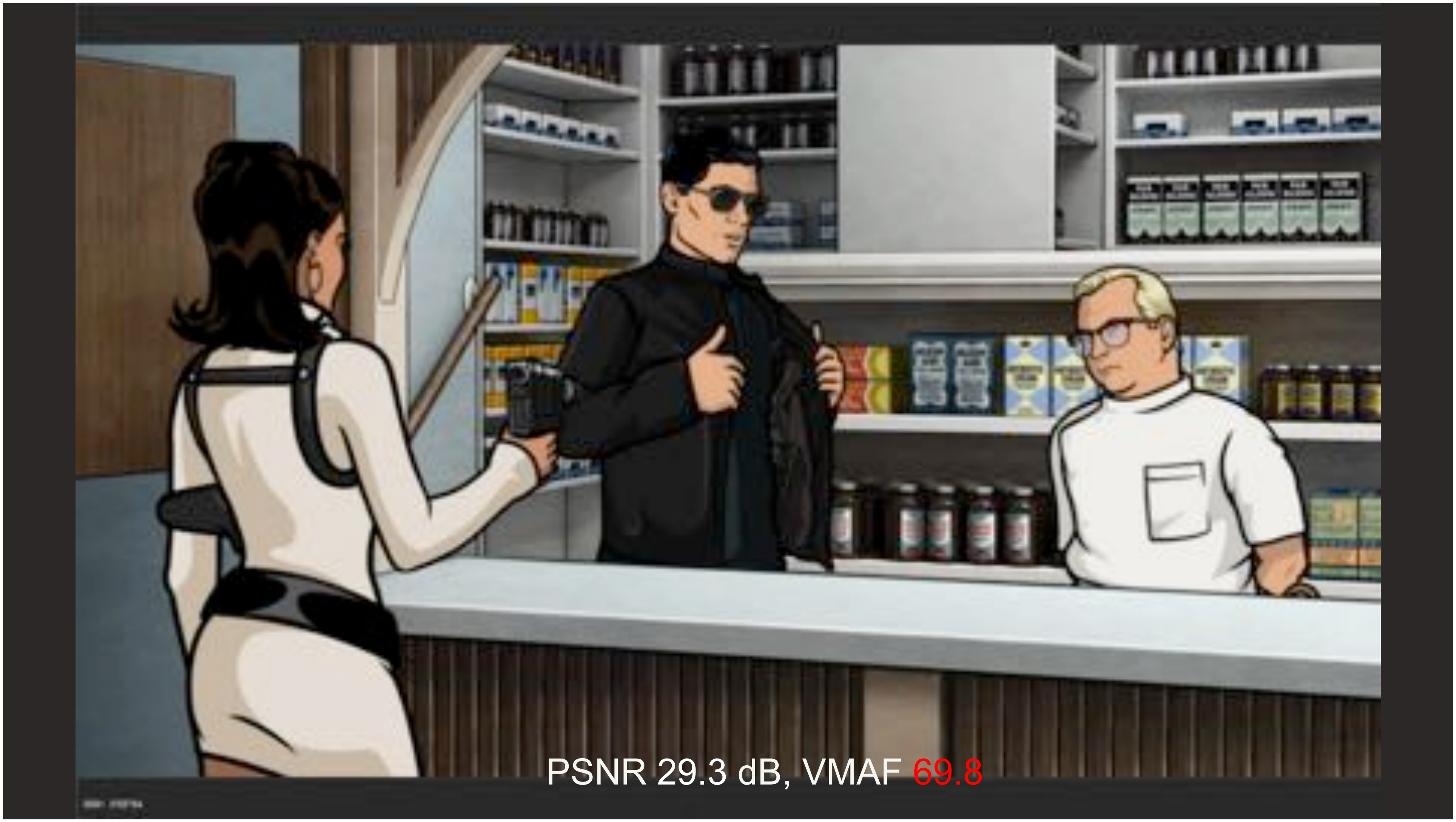
- Accurately measures **human perception** of quality
- Consistent across content
- Can be run at scale
- Works well relevant to adaptive streaming
  - Compression artifacts
  - Scaling artifacts

VMAF: Video Multimethod Assessment Fusion


PSNR 37.1 dB, VMAF 71.1


PSNR 32.9 dB, VMAF 70.2


PSNR 29.1 dB, VMAF 20.4


PSNR 29.3 dB, VMAF 69.8

Video Multimethod Assessment Fusion

- Full-reference video quality metric

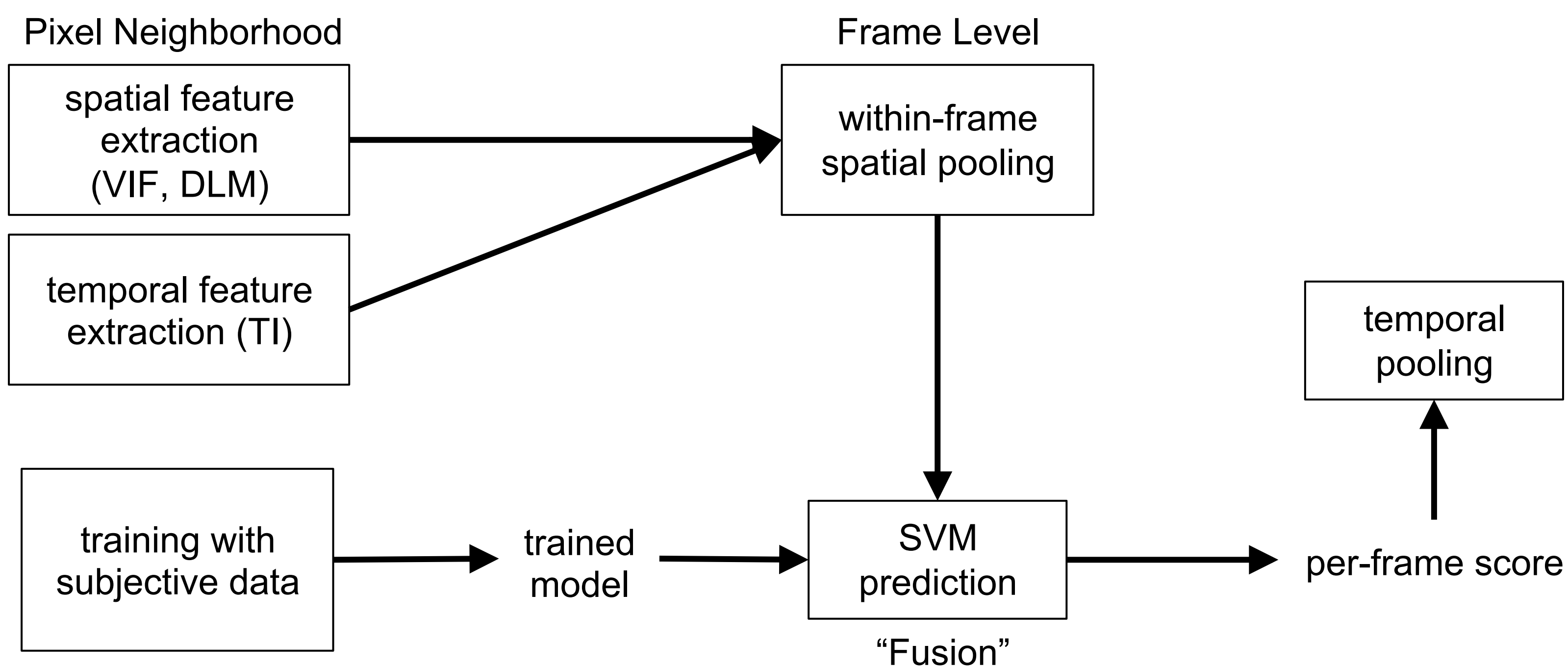
- Combines multiple elementary quality metrics
  - Visual quality fidelity (VIF*) @ 4 scales
  - Detail loss measure (DLM**)
  - Temporal information (TI) - average pixel difference between adj. frames

- Machine-learning regression to predict a final “fused” score, guided by subjective data

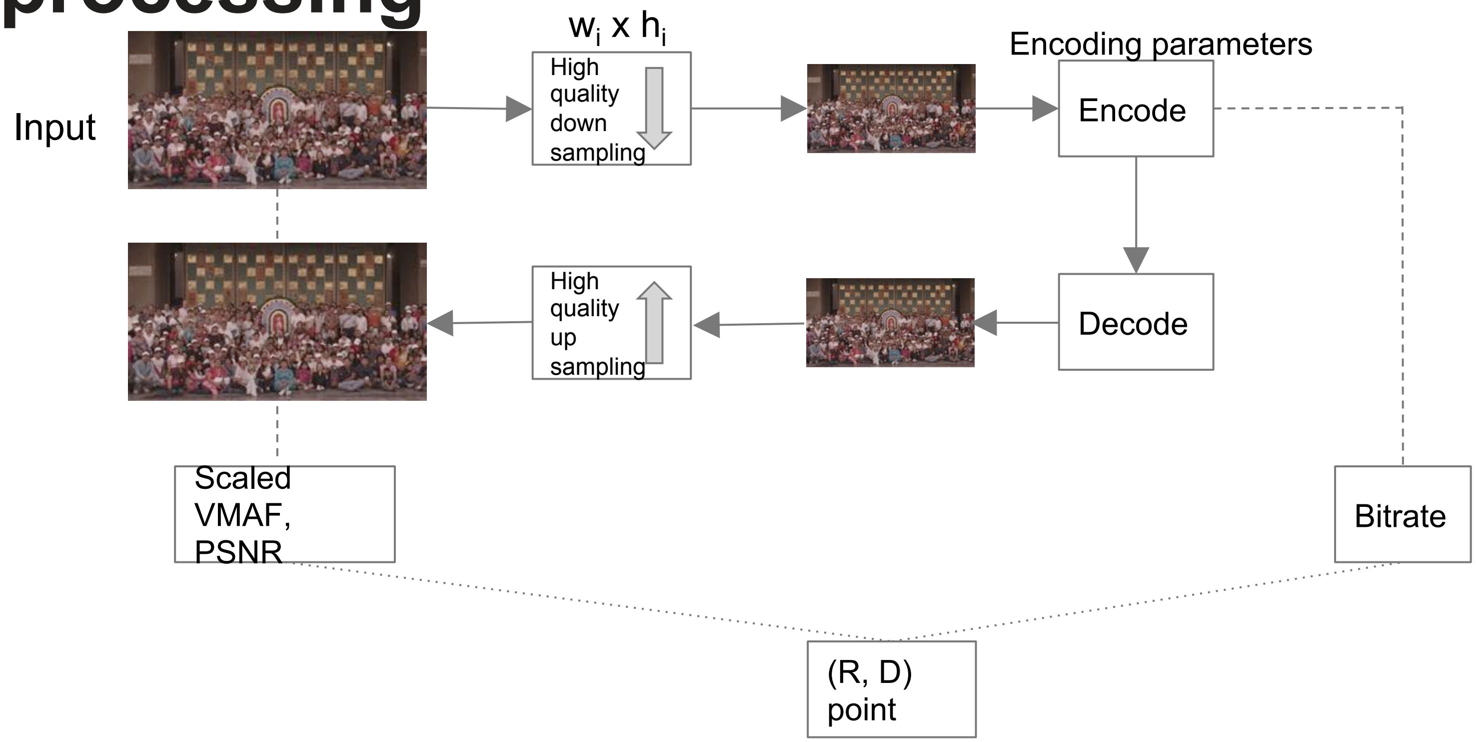
***Visual Information Fidelity -** H. Sheikh and A. Bovik, “Image Information and Visual Quality”.

****Detail Loss Measure -** S. Li, F. Zhang, L. Ma, and K. Ngan, “Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments”.

How VMAF works

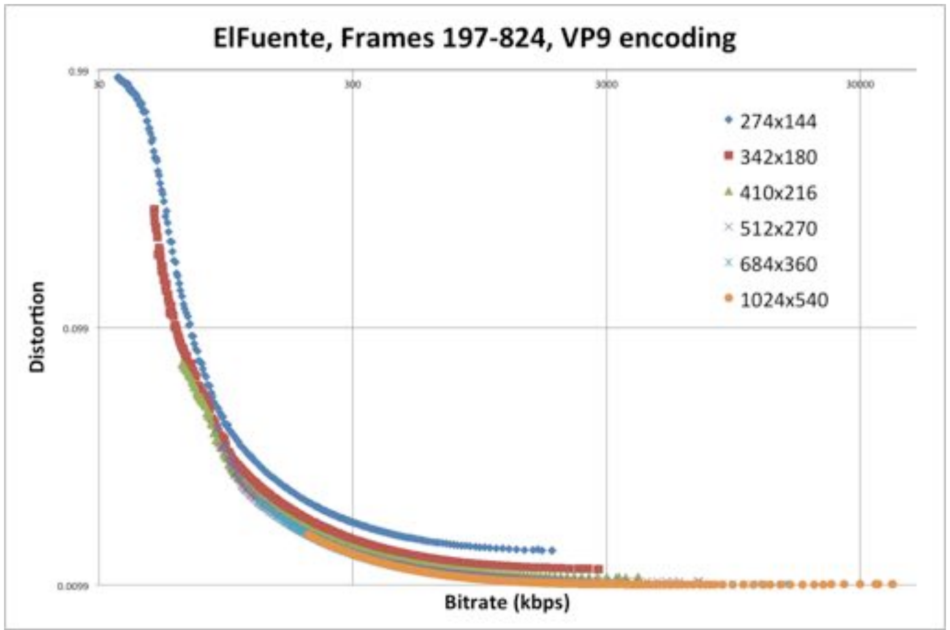


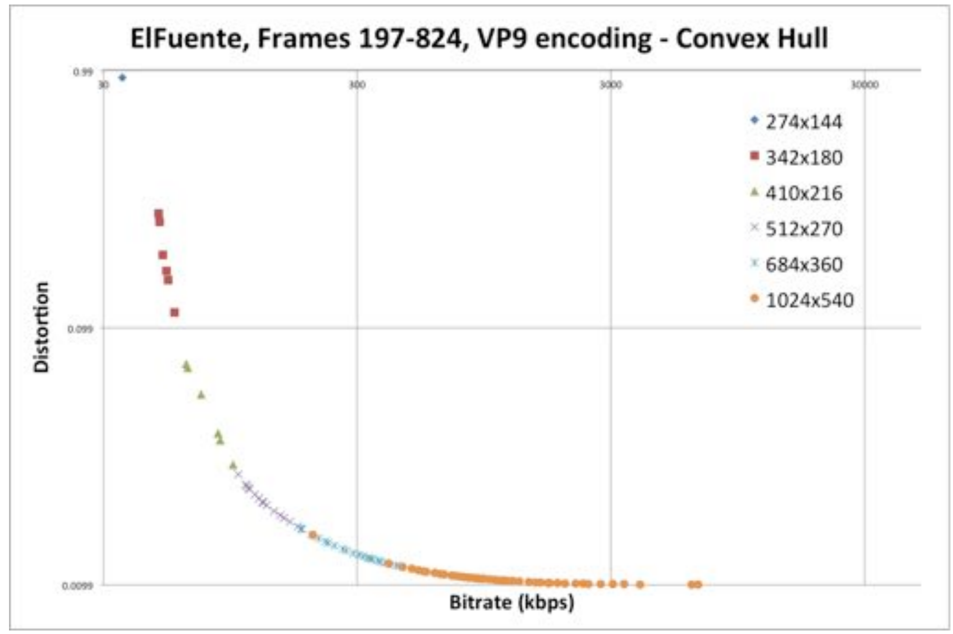
Dynamic Optimizer framework: Single shot processing



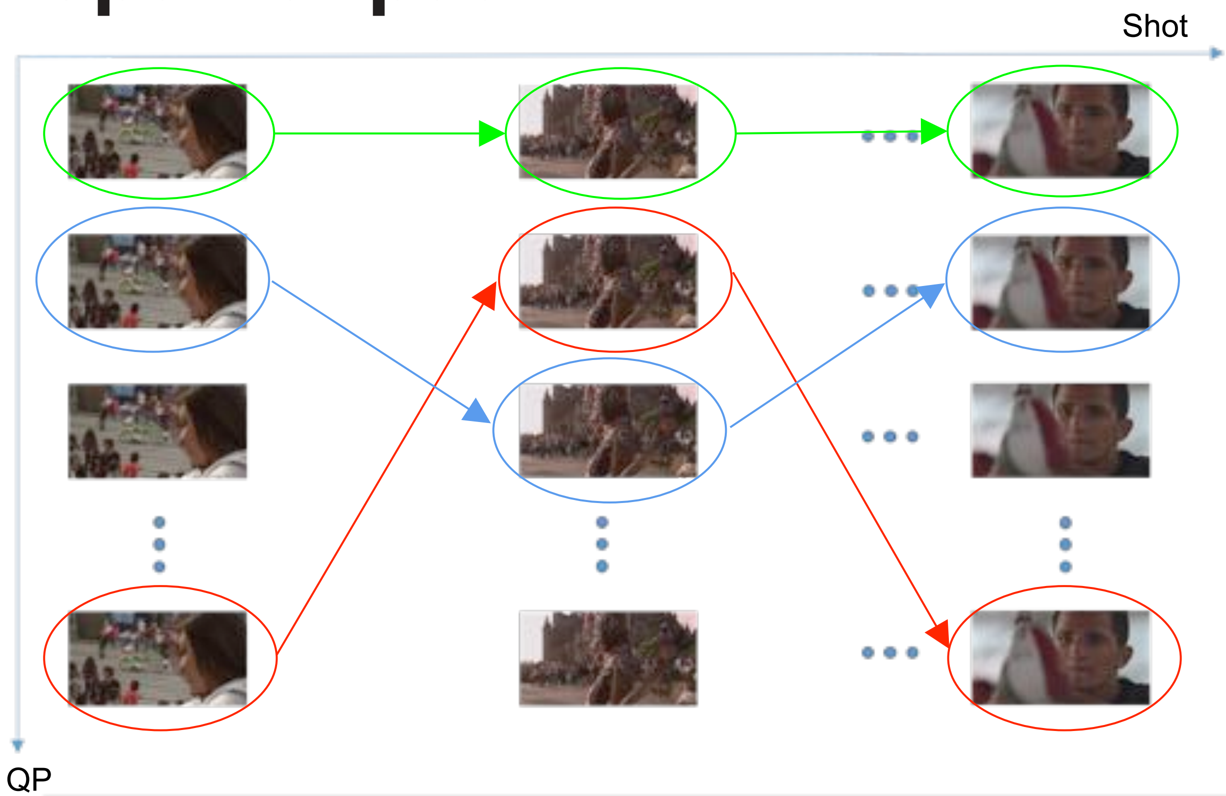
NETFLIX

Dynamic Optimizer framework: Convex hull of optimal shot encodes





Dynamic Optimizer framework: Trellis optimal path

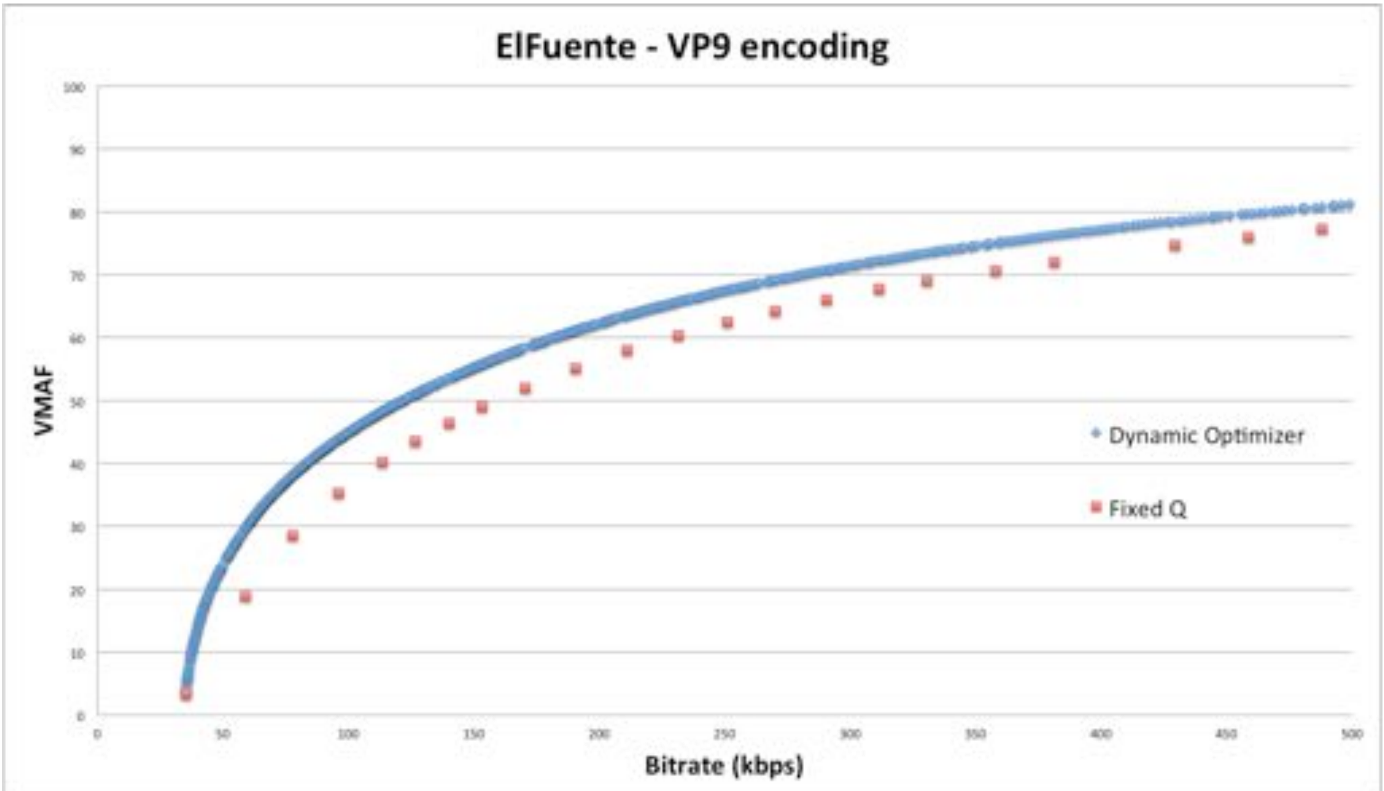


Fixed QP encode

Highest (average) quality encode, with bitrate x kbps

Lowest (average) bitrate encode, with quality y

Results: DO VP9 vs. Per-Title Optimal QP



NETFLIX

Video encoders used

- X264: open-source practical AVC encoder
- x265: open-source practical HEVC encoder
- EVE-VP9: practical VP9 encoder
- All in dual-tuning configuration

settings	x264	x265	EVE-VP9
profile	high	main	profile 0
preset	placebo	placebo	speed 1
number of titles	n/a	1	1
multi-threading	off	off	off
pass	1	1	1
PSNR-tuning	psy-rd=0	psy-rd=0 psy-rdoq=0	tune=psnr
perceptual-tuning	psy-rd=1.00	psy-rd=1.00 psy-rdoq=1.00	tune=visual

Quality metrics used

- TPSNR:

$$PSNR_{true} = 10\log_{10}(\frac{1}{N}\sum_{n=0}^{N-1}10^{-PSNRY_n} + 0.25\frac{1}{N}\sum_{n=0}^{N-1}10^{-PSNRCb_n} + 0.25\frac{1}{N}\sum_{n=0}^{N-1}10^{-PSNRCr_n})$$

- CPSNR:

$$PSNR_{classic} = \frac{1}{N}\sum_{n=0}^{N-1}PSNRY_n$$

- LVMAF:

$$VMAF_{linear} = \frac{1}{N}\sum_{n=0}^{N-1}VMAF_n$$

- HVMAF:

$$VMAF_{harmonic} = \frac{N}{\sum_{n=0}^{N-1}\frac{1}{1+VMAF_n}} - 1$$

Proposed codec testing methodology

- Use multiple encoding resolutions and QP values for each video sequence
- Use scaled metrics (PSNR, VMAF, ??) to construct R-D convex hull for each shot in a sequence
- Use dynamic optimizer to determine optimal encoding parameters for long sequences
- Use points equally spaced on the quality axis to calculate BD-rate
- Break down quality in ranges, when cross-over of R-D curves among different encoder configurations is observed


Breaking
Bad


NETFLIX
ORANGE
is the new BLACK


NETFLIX
HOUSE
of CARDS


MARVELS
AVENGERS

NETFLIX

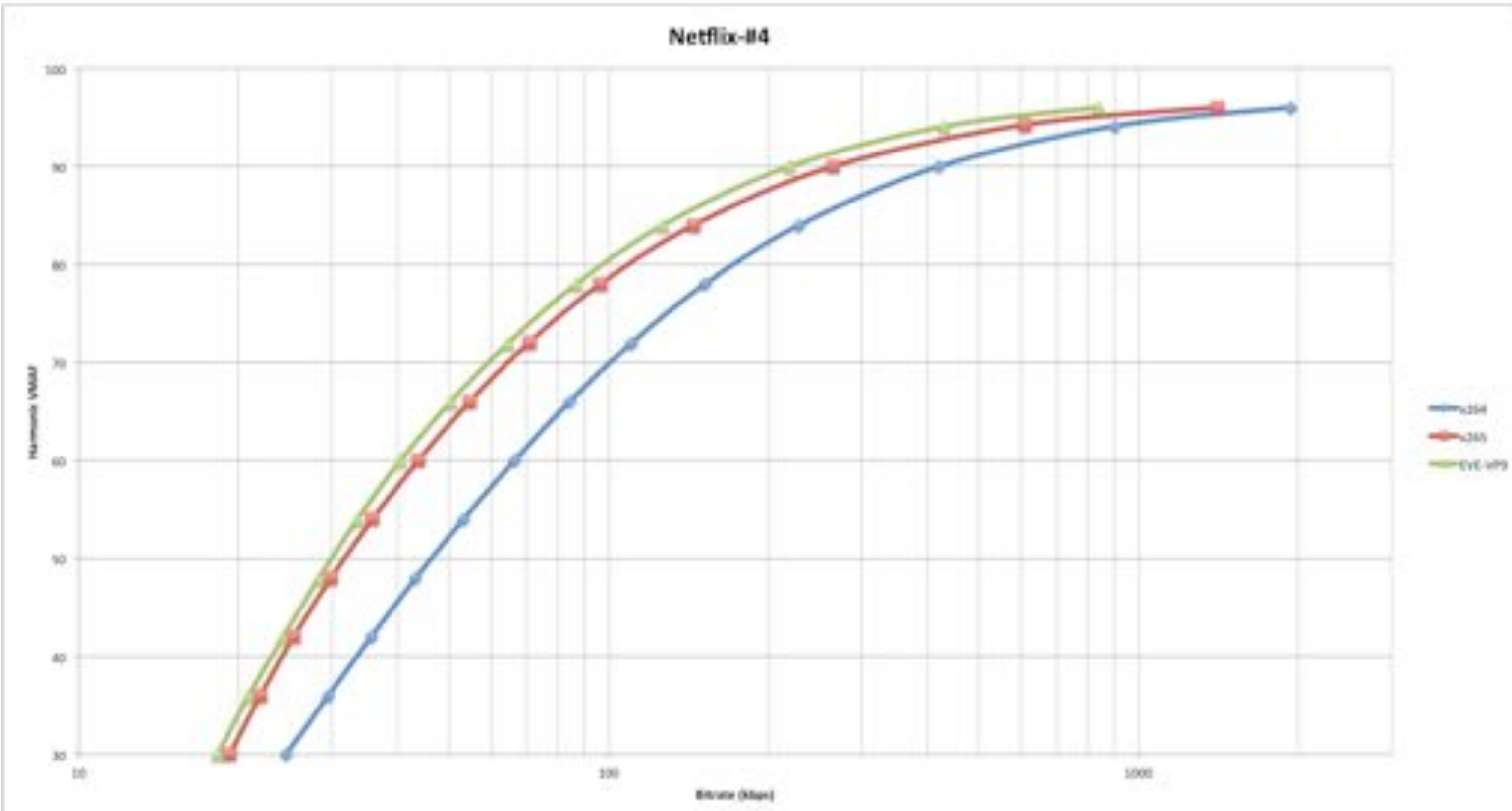

NETFLIX
BLOODLINE


NETFLIX
WET HOT
AMERICAN SUMMER
FIRST DAY OF CAMP

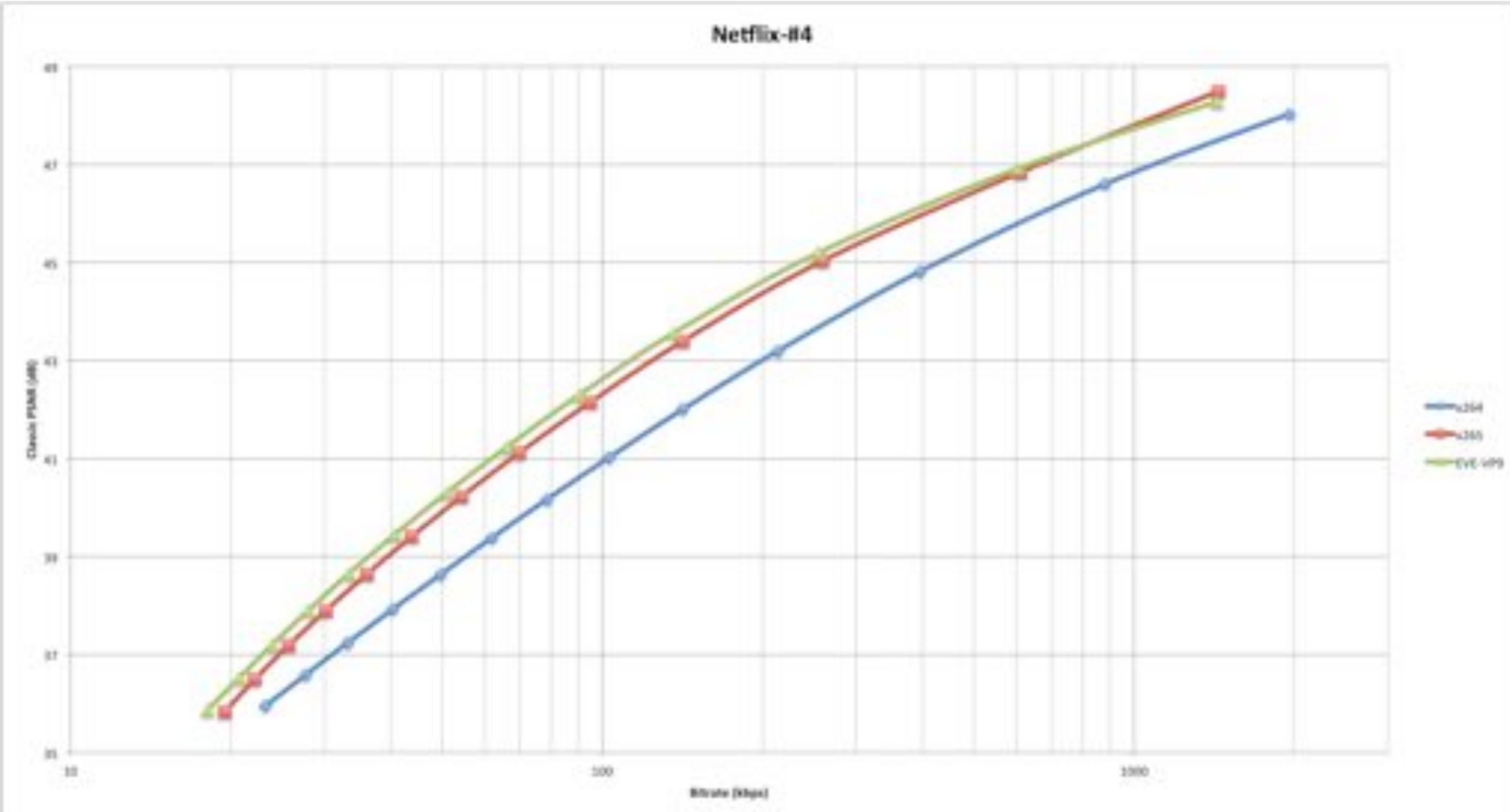

NETFLIX
MARVEL
DAREDEVIL


NETFLIX
BOJACK
HORSEMAN
HOLLYWOO

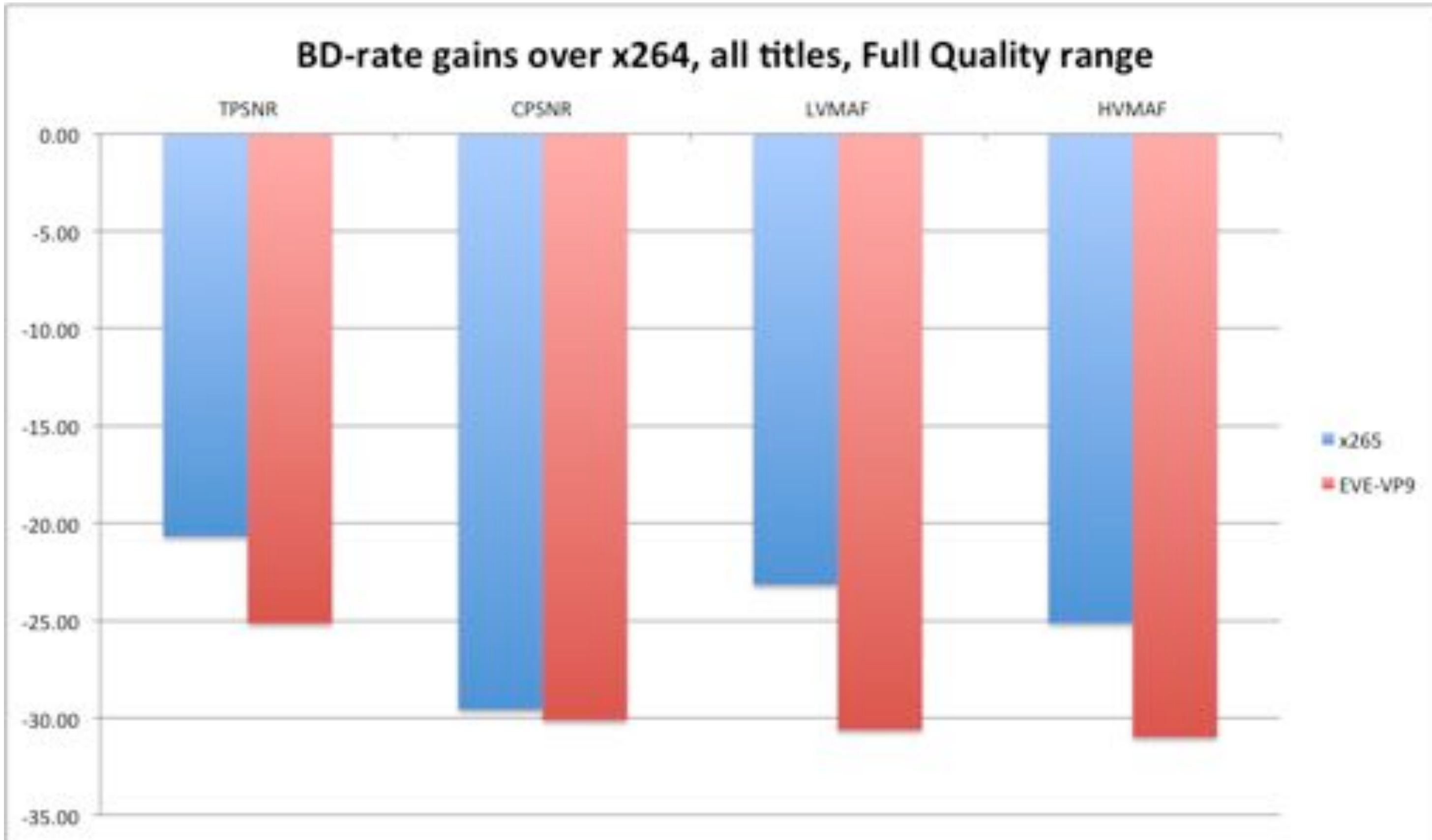
Results - 1



Results - 2

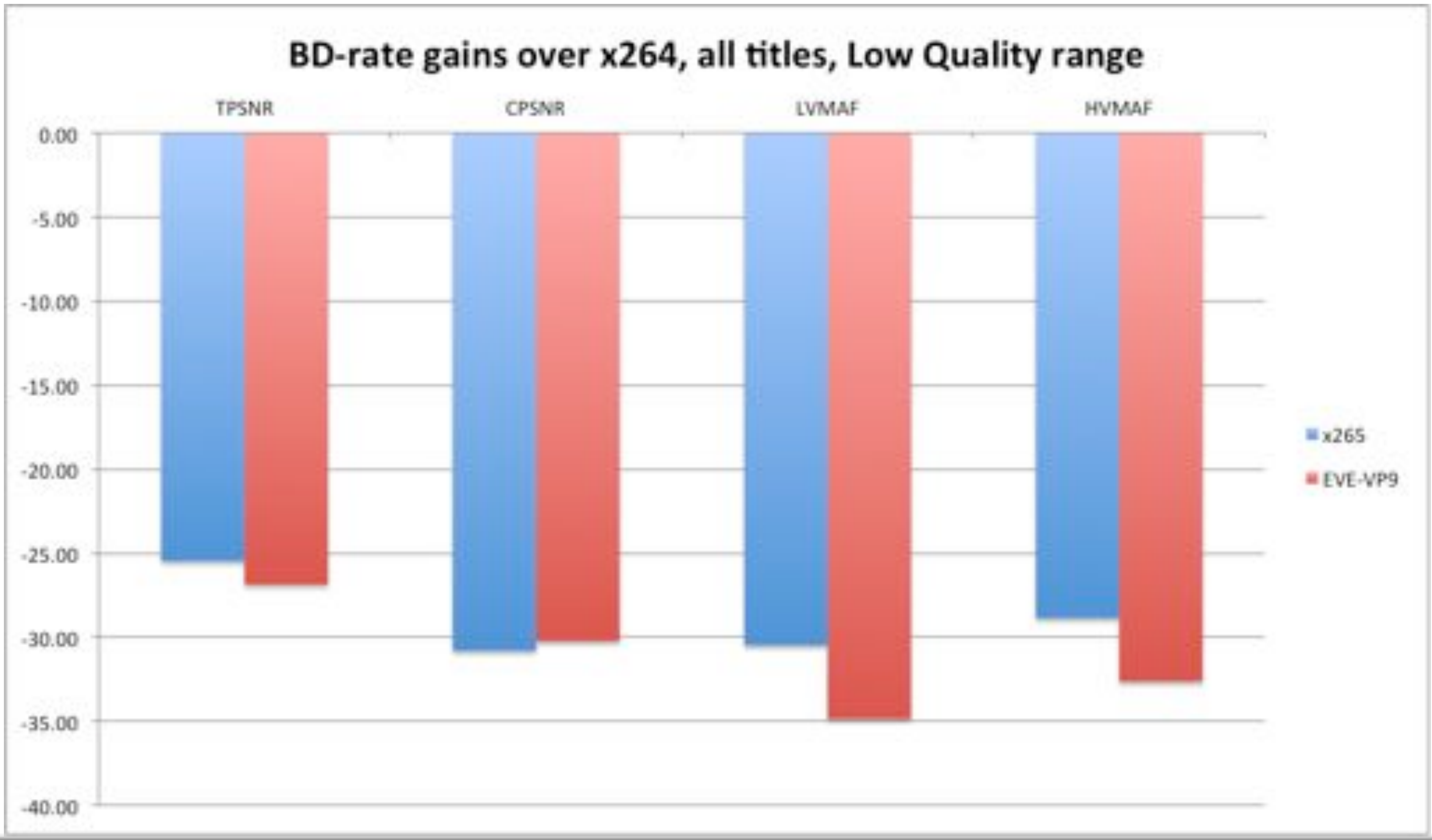


Results - 3



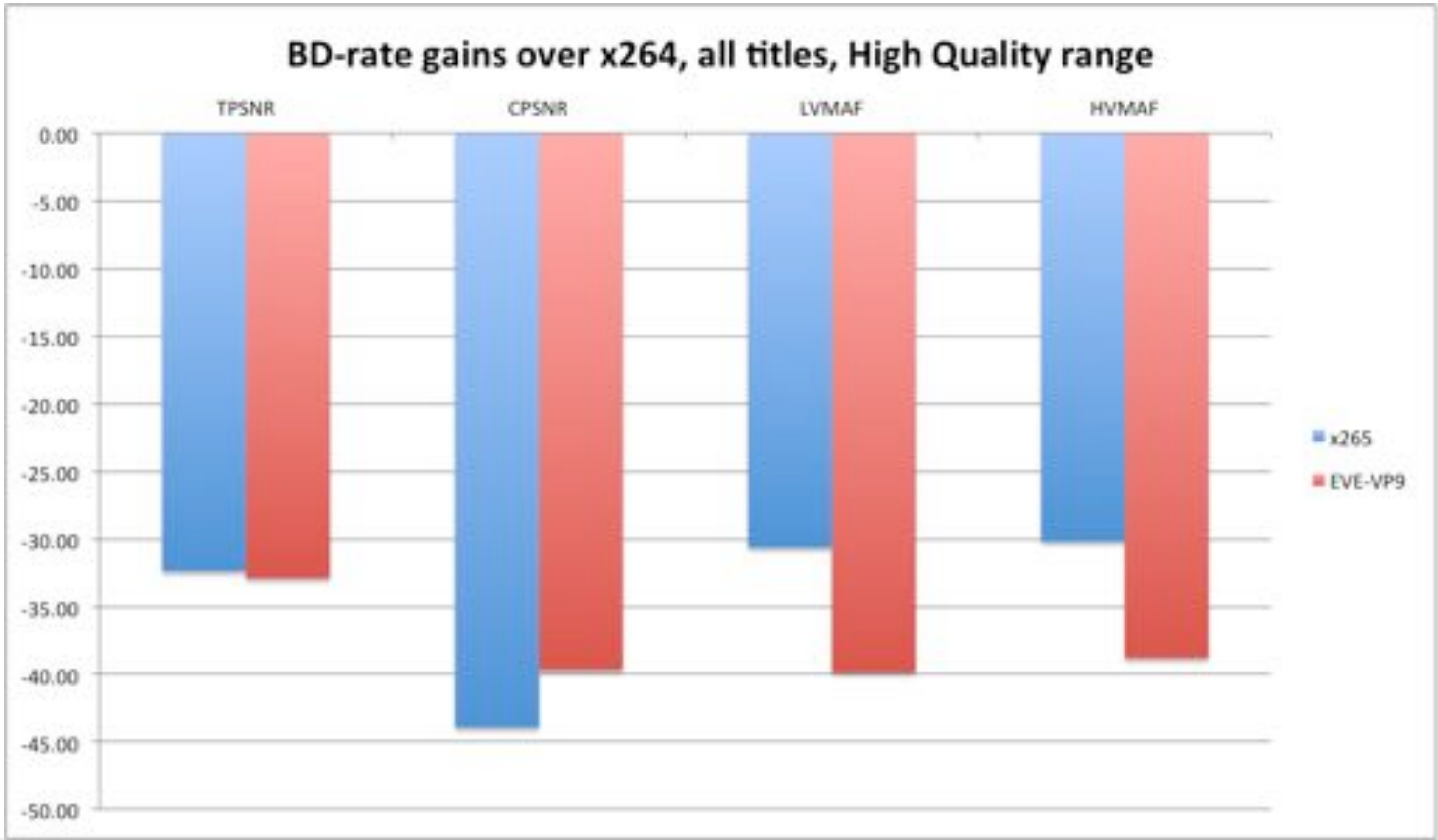
NETFLIX

Results - 4



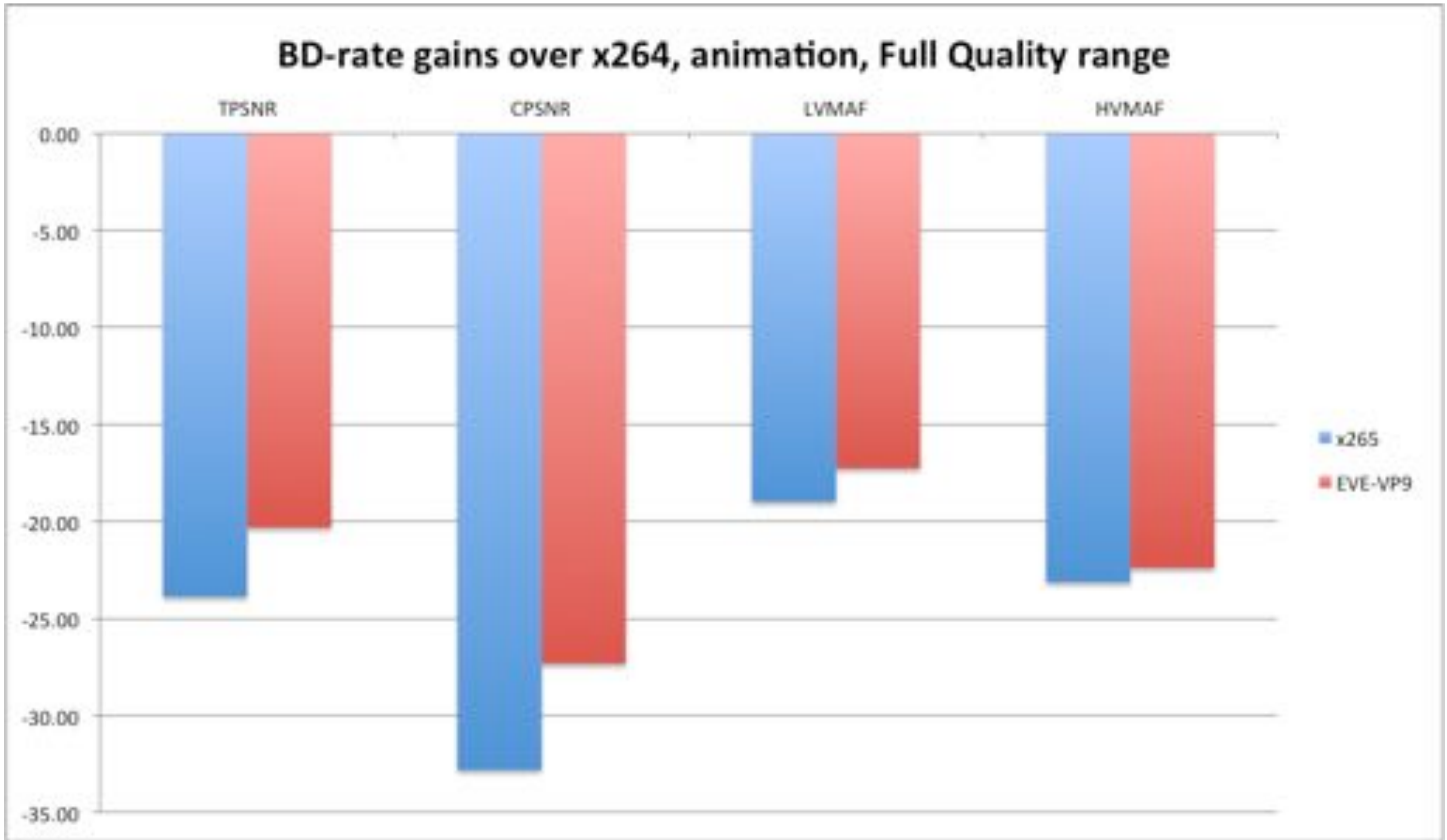
NETFLIX

Results - 5

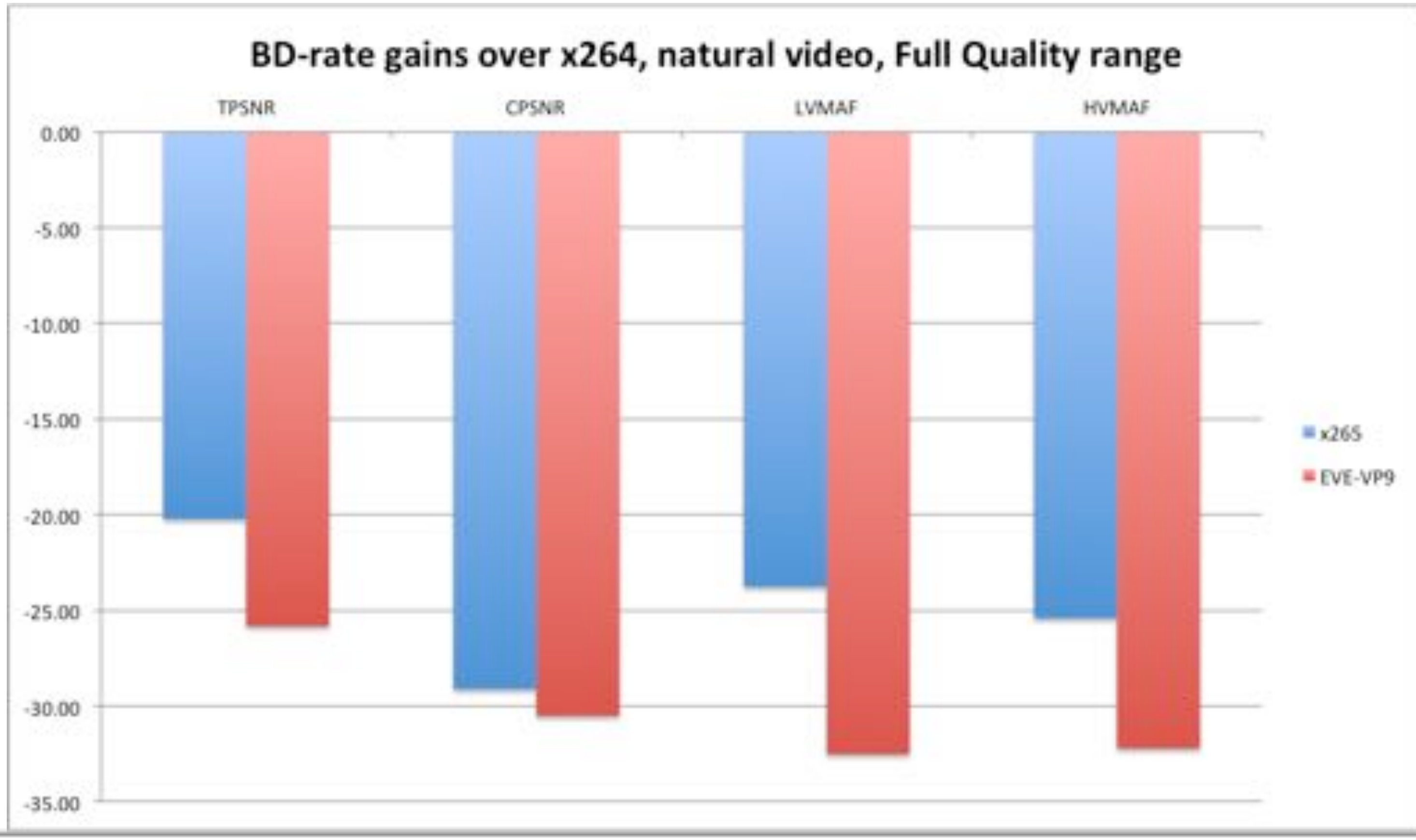


NETFLIX

Results - 6


BD-rate gains over x264, animation, Full Quality range
TPSNR
CPSNR
LVMAF
HVMAF
0.00
-5.00
-10.00
-15.00
-20.00
-25.00
-30.00
-35.00
x265
EVE-VP9

Results - 7



NETFLIX

A few facts

- x265 is about 20x slower than x264; EVE-VP9 is about 5x slower
- Test sequences represent about 8 hours of visual content
- Running these experiments took about 3 weeks to conclude
- 10M encodes produces
- 1M CPU hours – using cloud compute resources

Summary

- Scaled objective metrics better reflect adaptive streaming application video quality – VMAF more so than PSNR
- Using convex hull provides relevant QP ranges for each encoding resolution
- Dynamic optimizer offers an upper bound on RD-performance for long sequences
- DO allows for more fair comparison between codecs
- EVE-VP9 outperforms x265 for natural video sequences
- x265 outperforms EVE-VP9 for animation titles, mostly in low quality range

Video Algorithms Team @100M party

Academic research partners

University of Texas Austin

Prof. Al Bovik
Todd Goodall
Christos Bampis
Zeina Sinno

Université de Nantes

Prof. Patrick Le Callet
Lukáš Krasula

University of Southern California

Prof. C.-C. Jay Kuo
Joe Yuchieh Lin
Haiqiang Wang

University of Bristol

Prof. David Bull
Felix Mercer Moss
Mariana Afonso