Video Codec Testing and Quality Measurement

draft-daede-netvc-testing-01
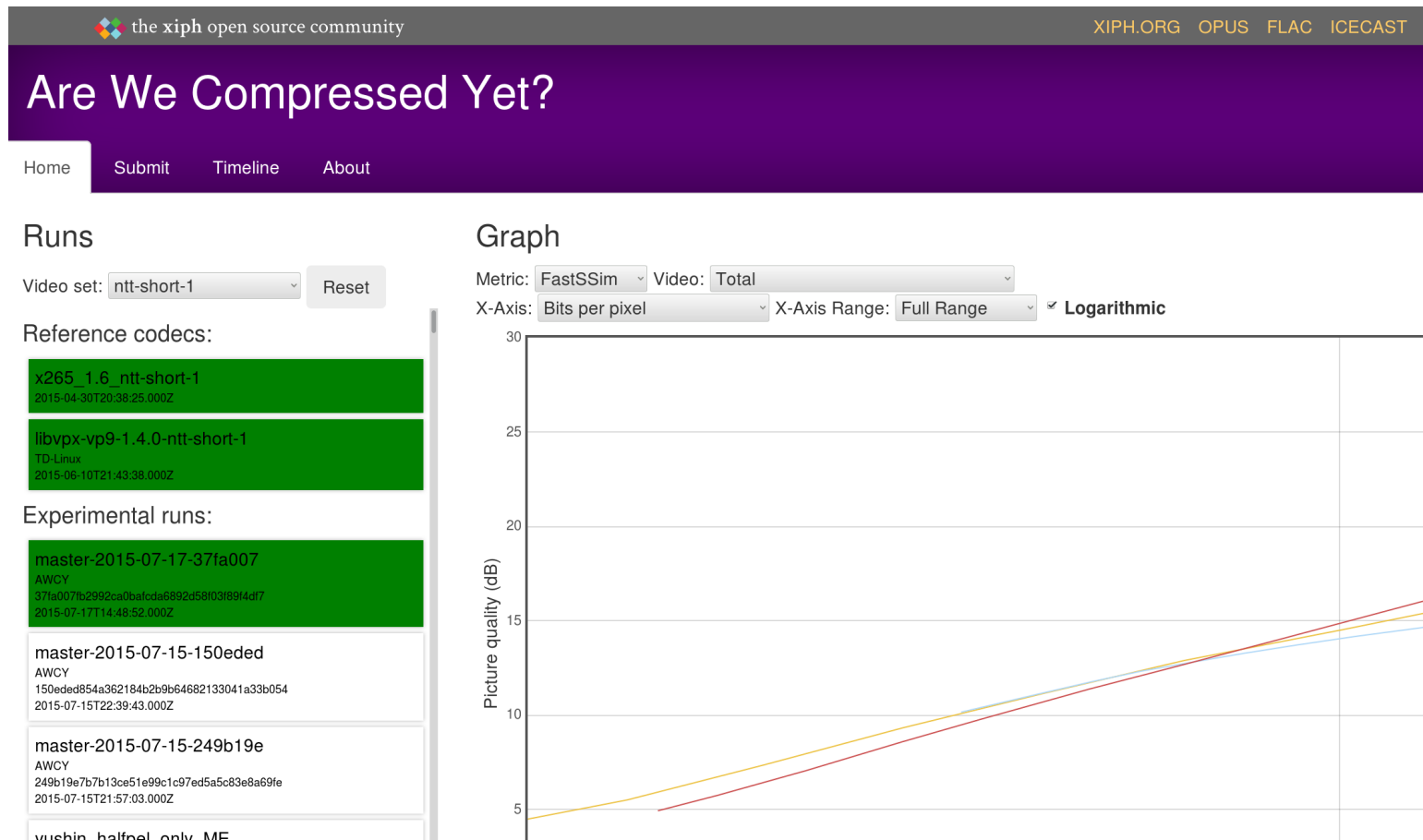
Thomas Daede

Overview

- Operating conditions of codec

- Method of quality measurement

- Testing material

AreWeCompressedYet?

- Testing described in this presentation is implemented at arewecompressedyet.com

Operating Parameter Choices

- Rate control method

- Maximum delay

- Frame type patterns

- Lookahead / 2 pass

JCT-L1100

- Specifies test sequences and operating conditions

  - Constant quantizer, no rate control

- Setup is HEVC specific

- Large source of contention for libvpx comparisons

Configuration: High-latency

- Test sequences are short (2-4 seconds, depending on frame rate), typical for DASH streaming or stored content

- All coding tools are allowed

- Long lookahead or two-pass is allowed

- Goal is to get maximum quality per filesize for test sequence

- Currently only supported mode in AWCY

Configuration: Low-latency mode

- Much longer sequences (15 seconds)

- No frame delay allowed

  - No lookahead

  - No out-of-order frames

- CBR mode with buffer model

Quality Measurement falls into two categories:

- Objective (metrics)

- Subjective (human testing)

Objective Metrics

- PSNR
  - Mean squared error on a log scale
  - Very easy to understand
  - Correlates poorly with visual quality

Without activity masking

10

With activity masking

-0.5dB PSNR

Objective Metrics

- SSIM
  - Common perceptual metric

- PSNR-HVS
  - PSNR in the frequency domain, taking the contrast sensitivity function (CSF) into account

- Fast MS-SSIM (FASTSSIM)
 - SSIM calculated at multiple scales to account for the CSF

Objective Metrics

- SSIM and FASTSSIM give linear results from 0-1, so the C implementations in the Daala codebase convert to a logarithmic decibel scale to match PSNR and PSNR-HVS
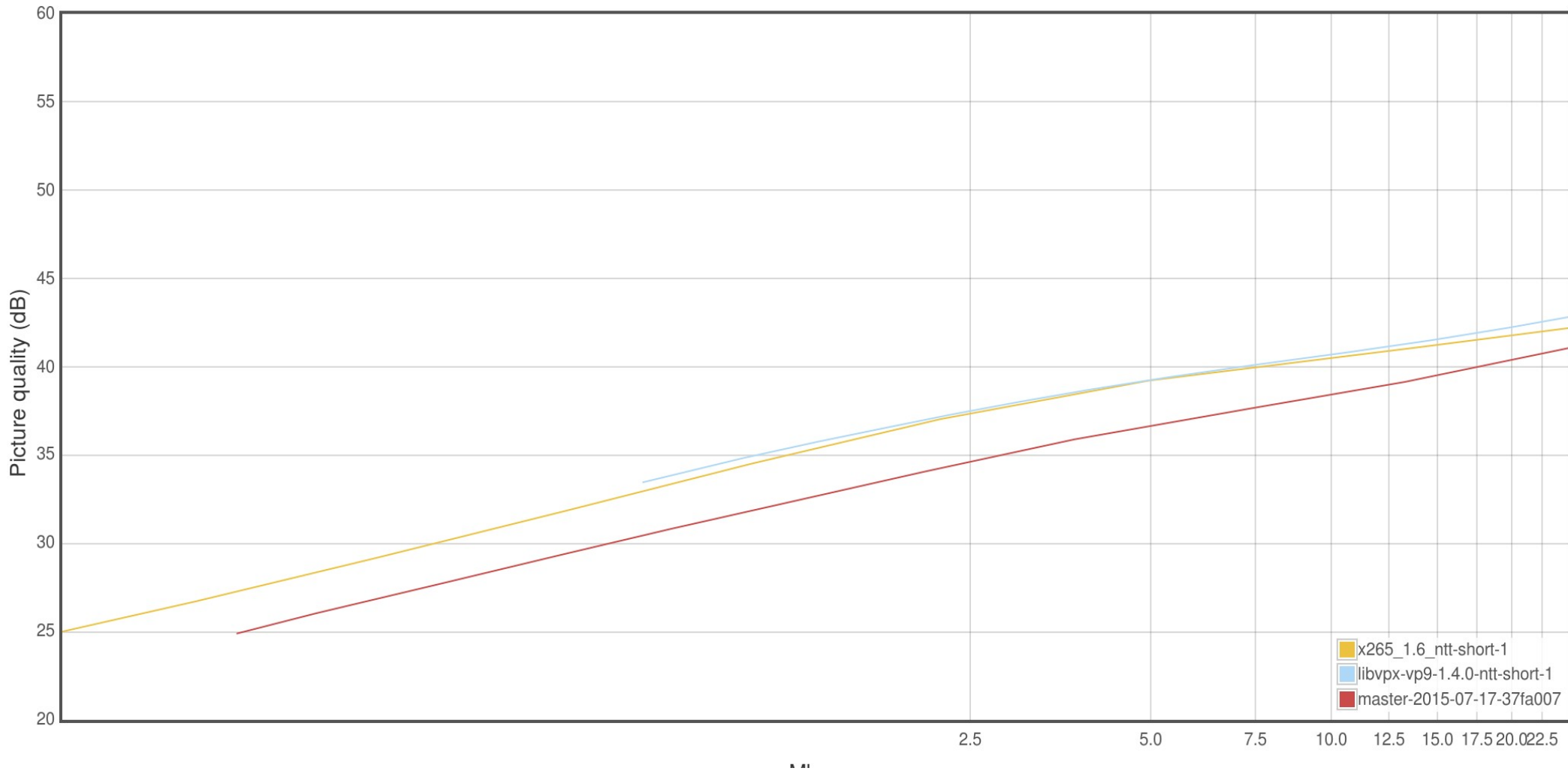
Objective Metric Pitfalls

- Only take into account luma

- Still image metrics
  - Input and output frames must have a 1:1 correspondence

- Implementations may give slightly different results
  - Use reference implementations in the Daala repository

Metric Graphing
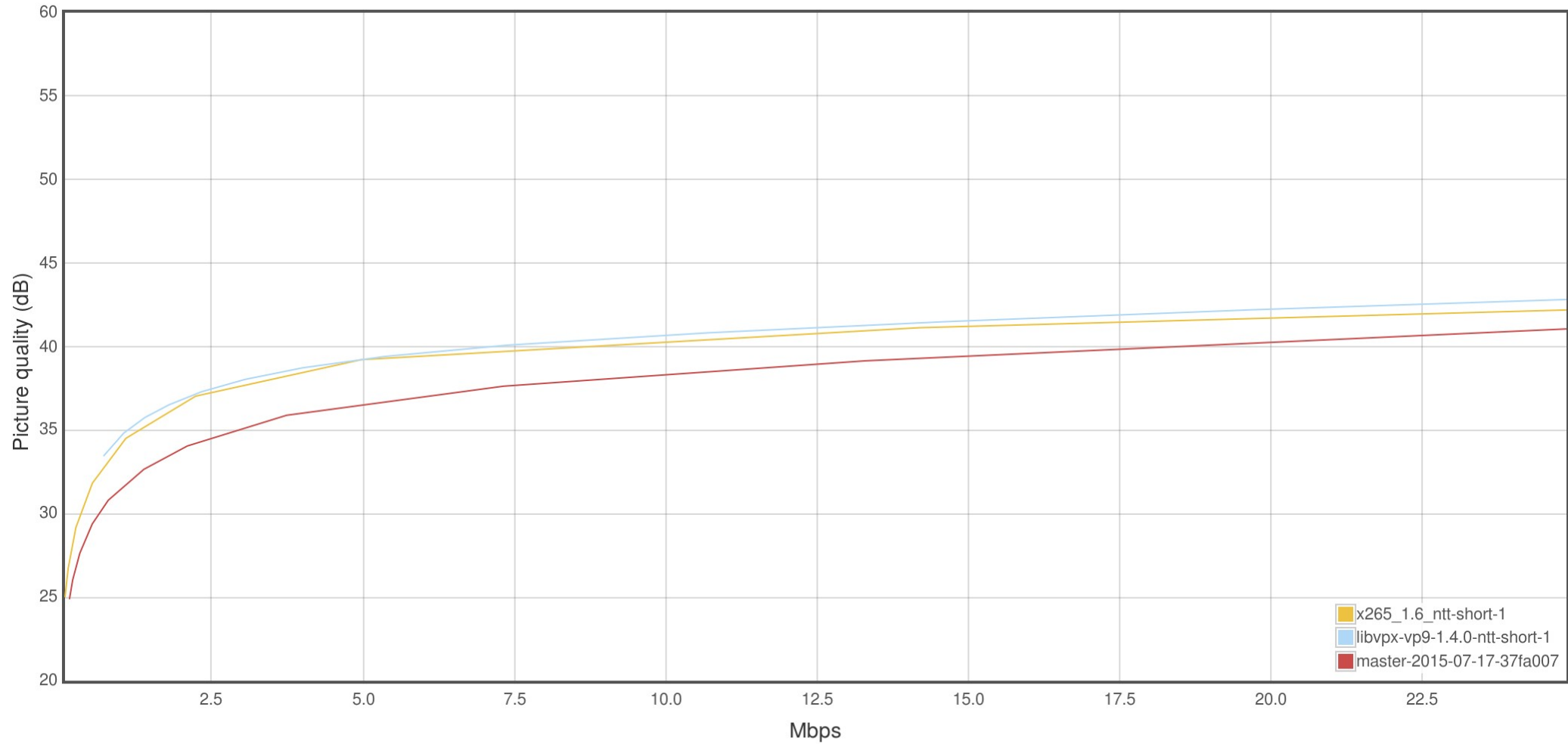
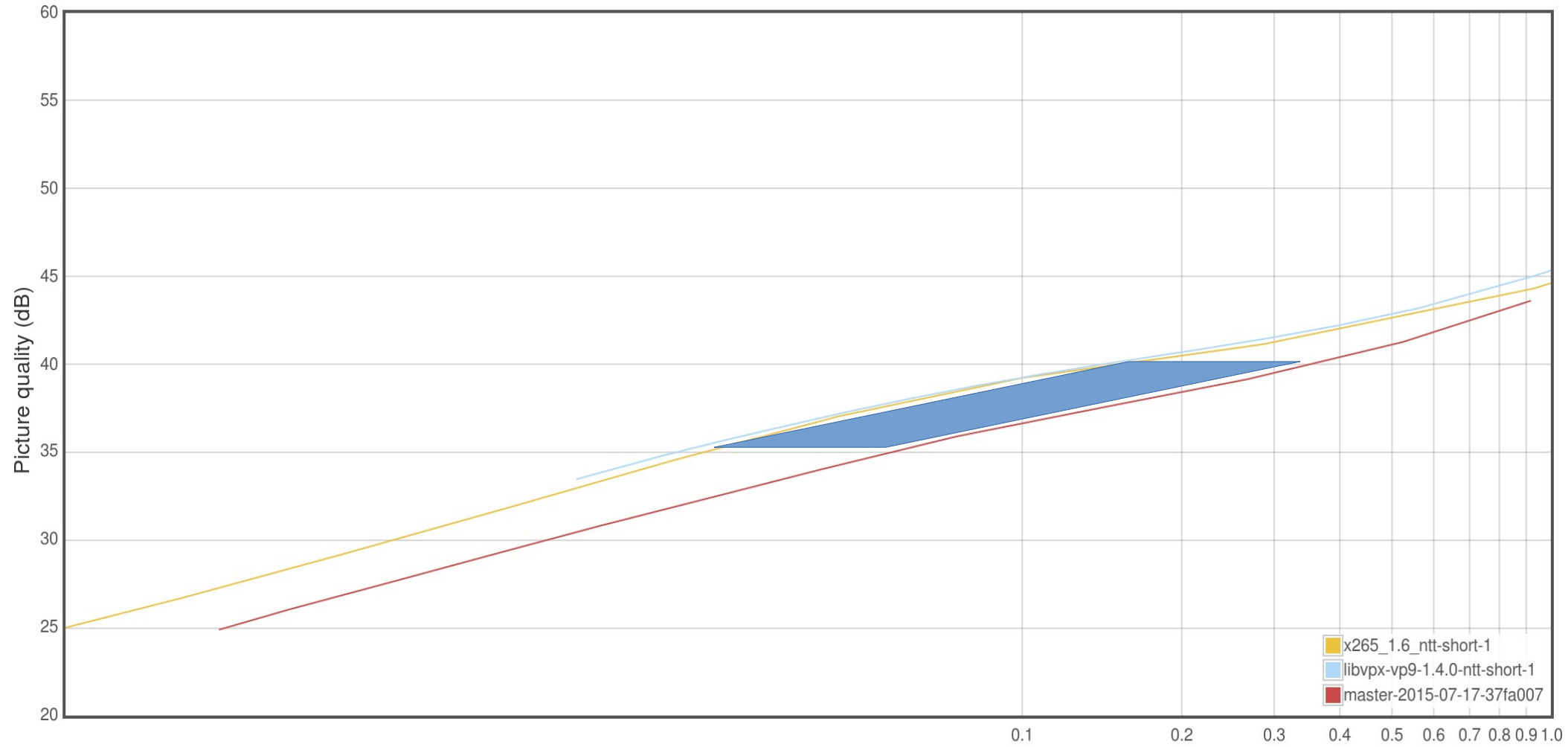## Linear

Metric Graphing

Logarithmic

BD-rate

- Computes the area difference between two curves in the log-log domain

- Gives a single number for a given range (e.g. codec A gives 50% the filesize for the same metric score as codec B)

BD-rate

Test clips

- Test clips organized into test sets

- Stored losslessly

- Preprocessing already applied (deinterlacing, frame dropping, etc)

- All clips in the set should have the same resolution and length

- Redistributable for replication of results

Test clips

- Video-hd-2

  – 1080p 4:2:0 clips from Derf's Test Video Collection (https://media.xiph.org/video/derf/)

- 4k Tears of Steel

- Screenshots

- Need more!

See More

- https://wiki.xiph.org/Daala

- Daala metric tools
  - https://git.xiph.org/?p=daala.git

- AreWeCompressedYet sources:
  - https://github.com/tdaede/rd_tool/
  - https://github.com/tdaede/awcy

Questions