

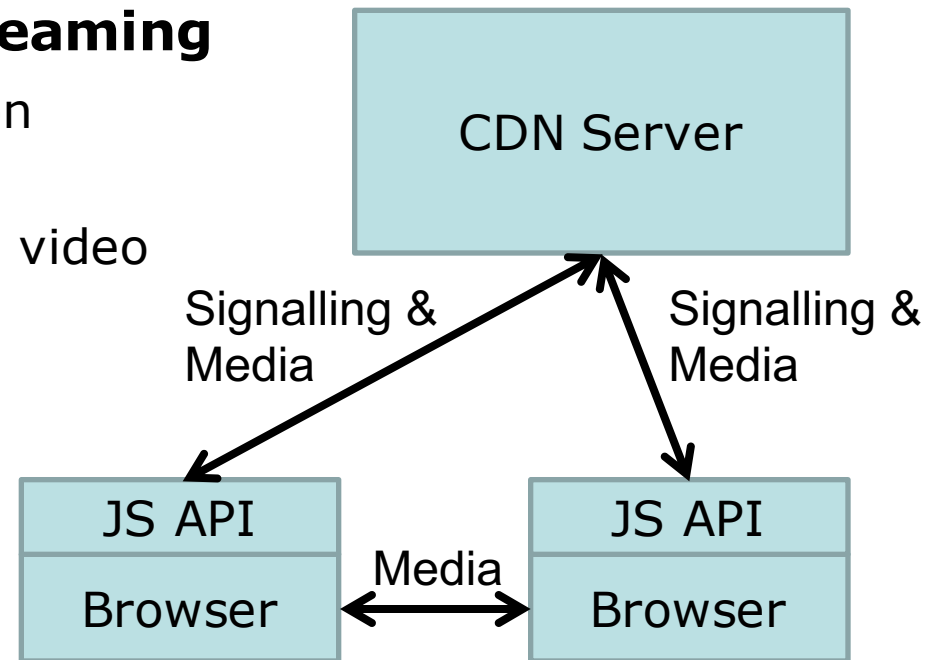
0. Motivation

WebRTC based distributed Streaming

- ❖ Browser-to-Browser communication framework
- ❖ Several start ups are investigating video content distribution with WebRTC
 - Including DASH, DRM, CDN
 - ~~Try webtorrent.io for demo~~

The question of incentive?

- ❖ *On the one hand:* much lower barrier to entry than in traditional P2P systems by just loading a website
- ❖ *On the other hand:* viewers provide their paid access capacity to cut bandwidth costs of content providers



1. Assumptions and Research Questions

This work assumes a consent-driven model ...

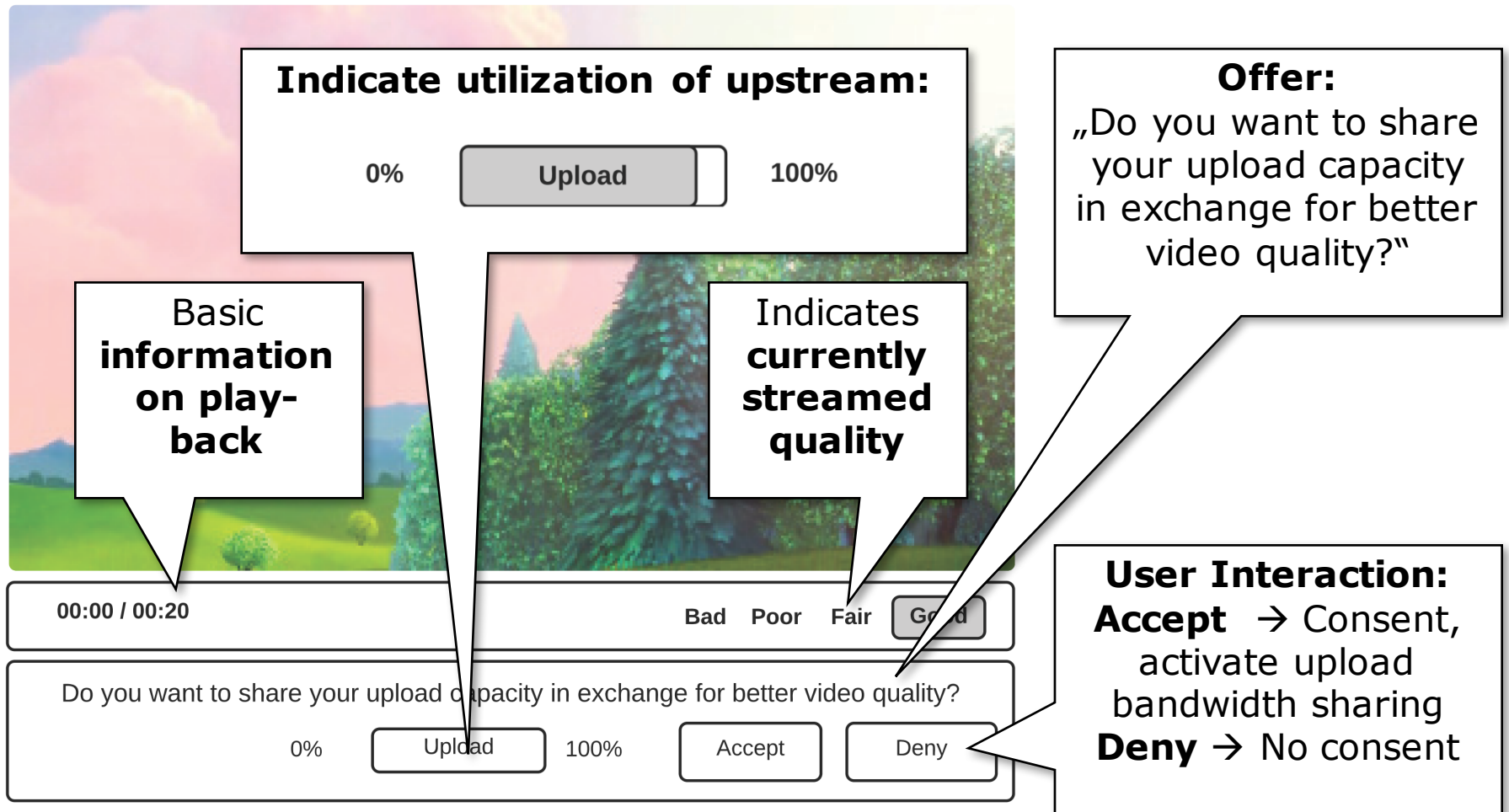
- Viewers are asked before upstream bandwidth is utilized
- An incentive can be provided by offering increased visual quality

Research questions ...

- a. How high is the **fraction of altruistic** users giving consent **without further benefits**?
- b. How high is the fraction of **non-altruistic users** that can be convinced to give their **consent in exchange for a better QoE**?
- c. Can incentive mechanisms based on **findings from behavioral economics be utilized** to increase consent?
- d. How sensitive are users to a **utilization of upload capacity without their consent**?

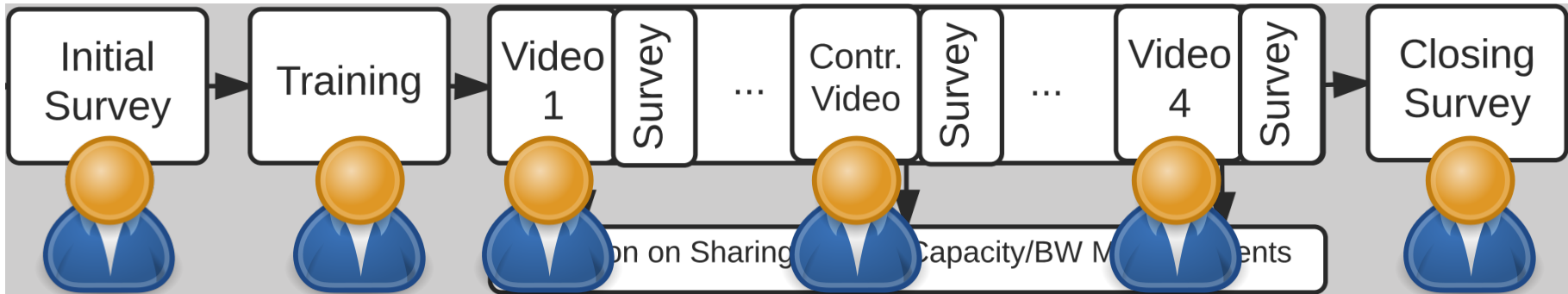
2. Crowdsourcing Study Design

User Interface



3. Experiment Structure

Subject Interaction with Experiment



- ❖ *Initial Survey:* demographics of the sample
- ❖ *Training:* visual quality training, upload capacity training
- ❖ *Video X:* subject watches video and receives offer, can accept or deny
- ❖ *Interm. Survey:* capture motivation of decision
- ❖ *Closing Survey:* general questions

4. Endowment and Control Treatment

Endowment Treatment (50% of subjects)

- ❖ Beh. Economics: **People attribute a higher value to goods merely because they are owned due to loss aversion (Endowment Effect).**
 - Mr. R buys a case of good wine \$5 a bottle
 - Later, his wine merchant offers to buy the wine back for \$100 a bottle
 - Mr. R refuses, although he has never paid more than \$35 for wine.

Idea - Create feeling of loss:

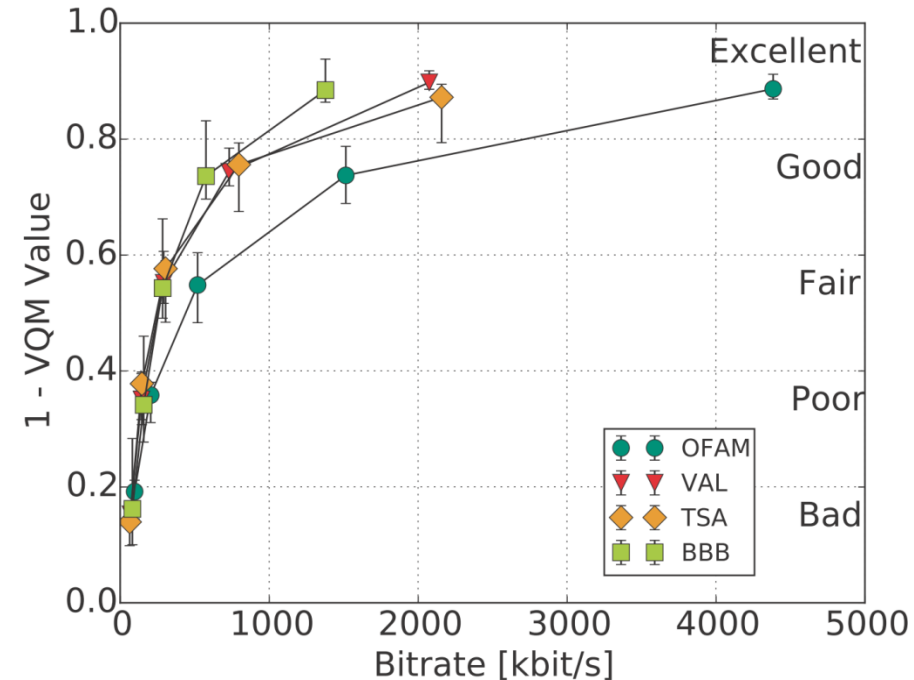
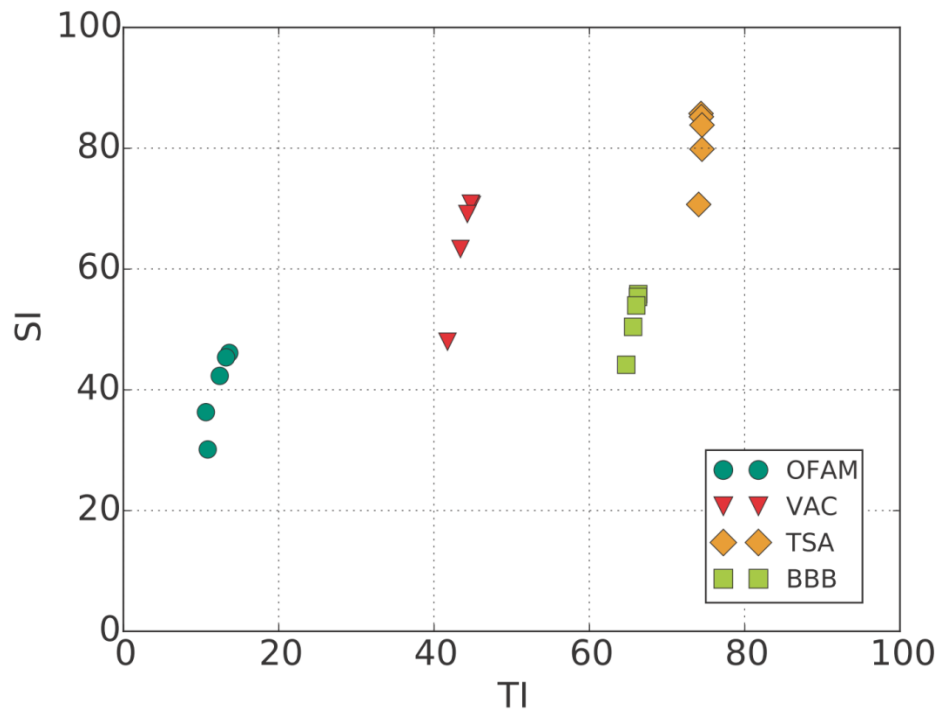
show 10s high video quality first → downgrade → place offer

Control Treatment (50% of subjects)

Avoid feeling of loss:

show 10s low video quality first → place offer

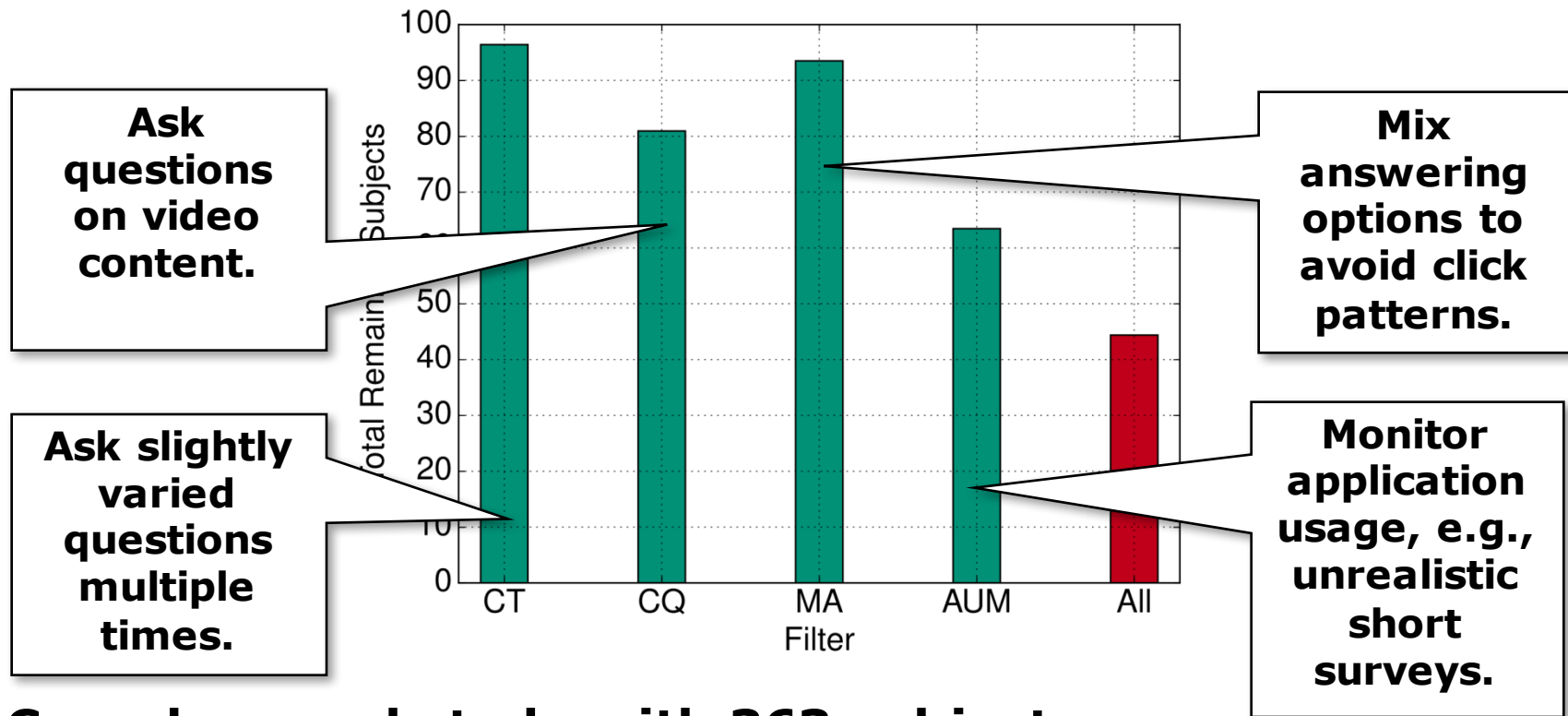
5. Video Sequence Preparation



❖ Lederer et al.'s DASH test set [19]

- Covers TI/SI space
- Reencoded to have similar VQM [16] distances between layers

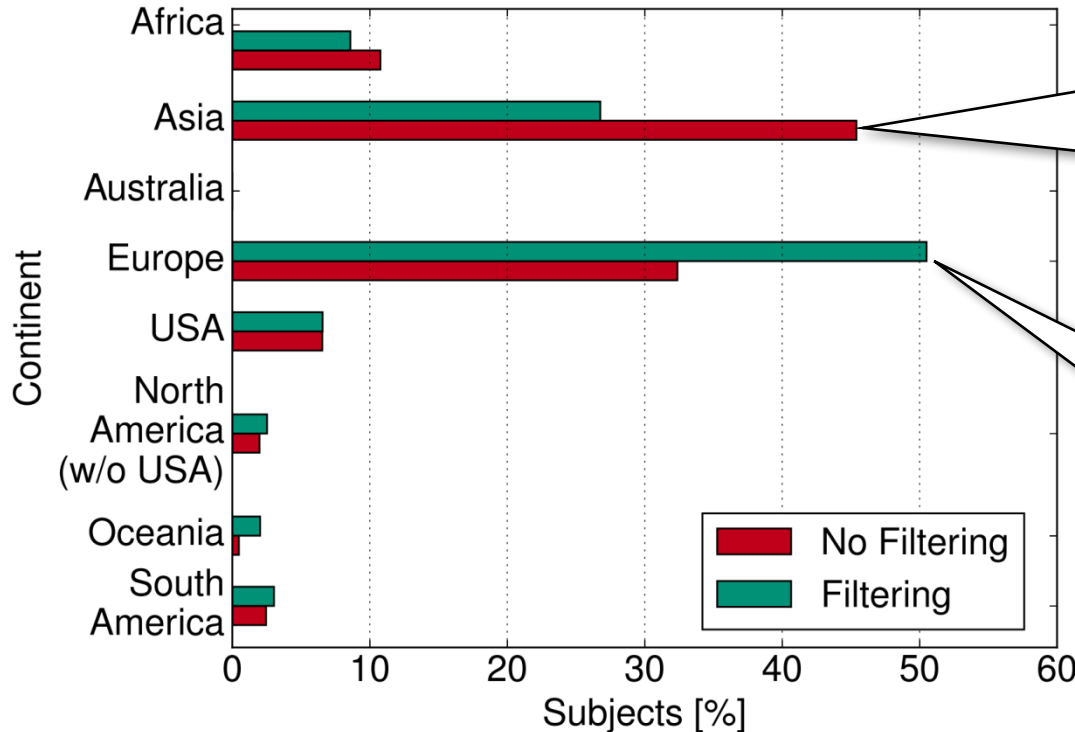
6. Reliability Filtering



Crowdsourced study with 363 subjects

- ❖ Microworkers, 0.4\$ per sample
- ❖ 192 subjects remain after reliability filtering

7. Sample Description



Interesting effect:

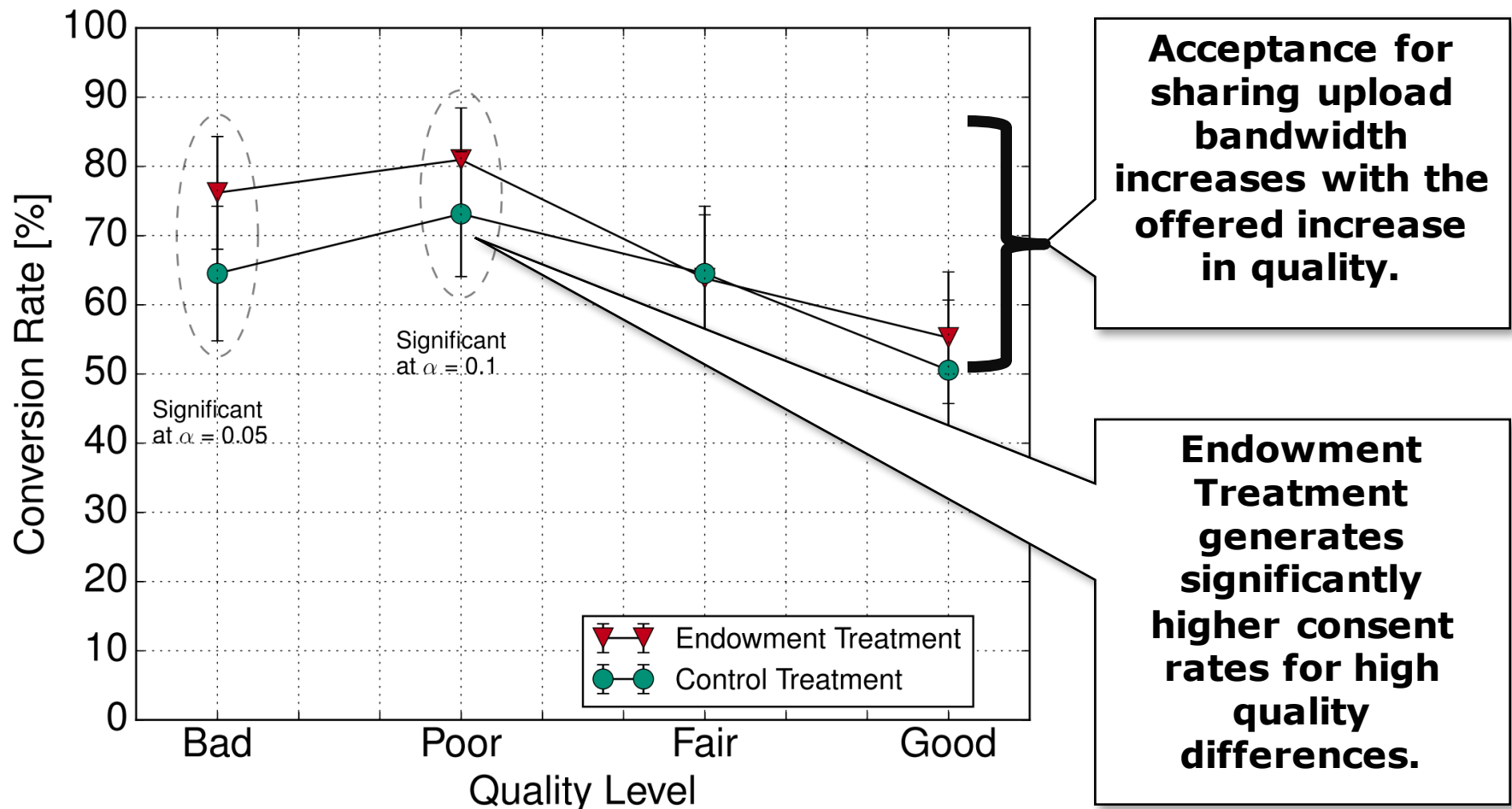
Asian subjects were likely to be filtered.

Majority of subjects are from Europe.

International young sample

- ❖ Median age between **26 and 30**
- ❖ Subjects state they are affine to digital technology

8. Results Endowment vs. Control



9. Group Structure

Bandwidth Agnostic

- ❖ **No valuation for upstream bandwidth**, always accept offer even when no quality increase is offered

Quality Aware

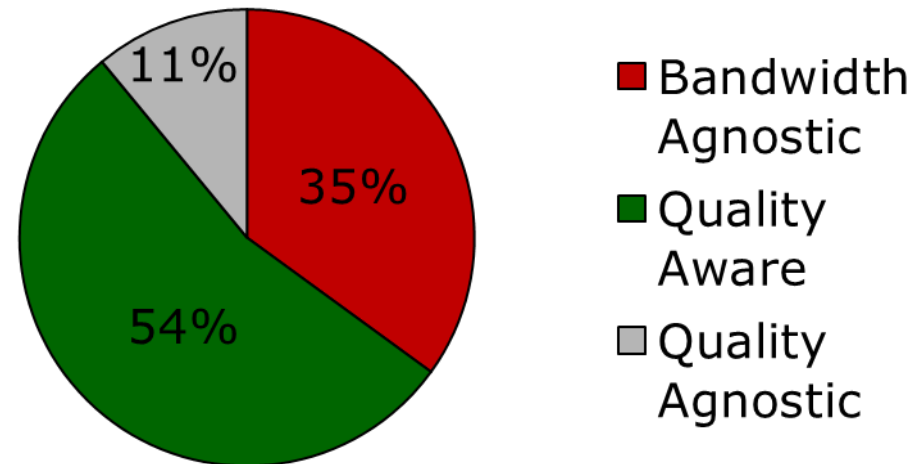
- ❖ **Decision based on the offered increase in quality**

Quality Agnostic

- ❖ **No valuation for quality, always deny offer.**

→ For motivation of behavior, please see paper.

User Group Share



Quality Aware and Quality Agnostic Subjects (**65%**) think **utilization of upstream capacity without consent is unacceptable.**

10. Summary

- a. How high is the **fraction of altruistic** users giving consent **without further benefits**?
 - 35% in our sample.
- b. How high is the fraction of **non-altruistic users** that can be convinced to give their **consent in exchange for a better QoE**?
 - 54% in our sample, the higher the change in quality offered, the more likely the group's consent.
- c. Can incentive mechanisms based on **findings from behavioral economics be utilized** to increase consent?
 - Utilizing the endowment effect increases consent rates by up to 12% compared to a control treatment.
- d. How sensitive are users to a **utilization of upload capacity without user's consent**?
 - 65% think it is unacceptable

11. Outlook

Implications for WebRTC video distribution

- ❖ **Content providers should always ask for user's consent** before leveraging WebRTC for content distribution
- ❖ **Offering increased QoE is useful** to increase consent rates
- ❖ The **endowment effect can be utilized** to maximize consent

Outlook

- ❖ Long-term stability of consent rates?
- ❖ Implications on streaming system architecture?
- ❖ Can the group structure be utilized (altruistic subjects)?
- ❖ Does utilizing the endowment effect have a positive effect on offloading server infrastructure?

12. Discussion

❖ Thank you for your attention

RT-VQM: Real-Time Video Quality Assessment for Adaptive Video Streaming Using GPUs



TECHNISCHE
UNIVERSITÄT
DARMSTADT

ACM Multimedia Systems Conference (ACM MMSys)
May 10-13, Klagenfurt, Austria



Dipl.-Wirtsch.-Inform. **Matthias Wichtlhuber**
Gregor Wicklein, M. Sc.
Stefan Wilk, M. Sc.
Prof. Dr. Wolfgang Effelsberg
Prof. Dr. David Hausheer

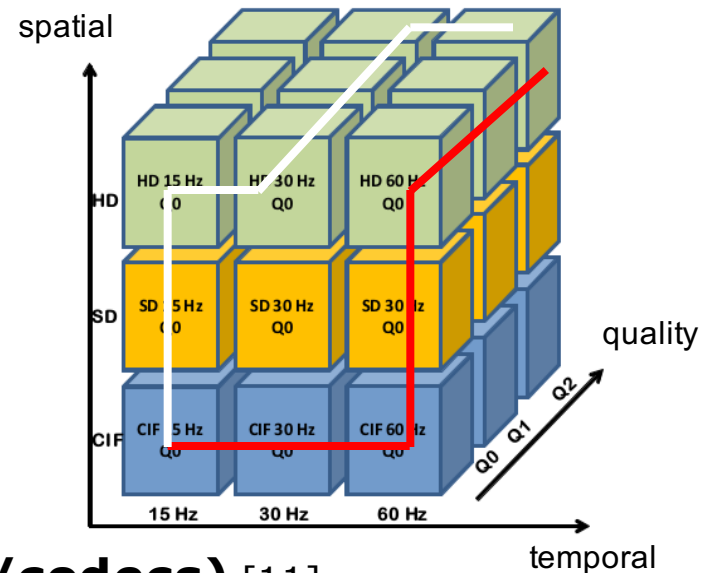
Visual Quality
Impairment?

PS – Peer-to-Peer Systems Engineering
Prof Dr. David Hausheer
Dept. of Electrical Engineering and Information Technology
TUD – Technische Universität Darmstadt
Rundeturmstr. 10, D-64283 Darmstadt, Germany
Tel. +49 6151 166150, Fax. +49 6151 166152
www.PS.tu-darmstadt.de

0. Motivation

Increasing relevance of live streaming use cases

- ❖ User generated content: Meerkat, Periscope, Twitch.tv
 - Twitch is 4th largest traffic source in the US, steady 8% growth rate [12]
- ❖ Telepresence, video conferencing, upcoming WebRTC standard



(#layers DASH/SVC) x (#devices/codecs) [11]

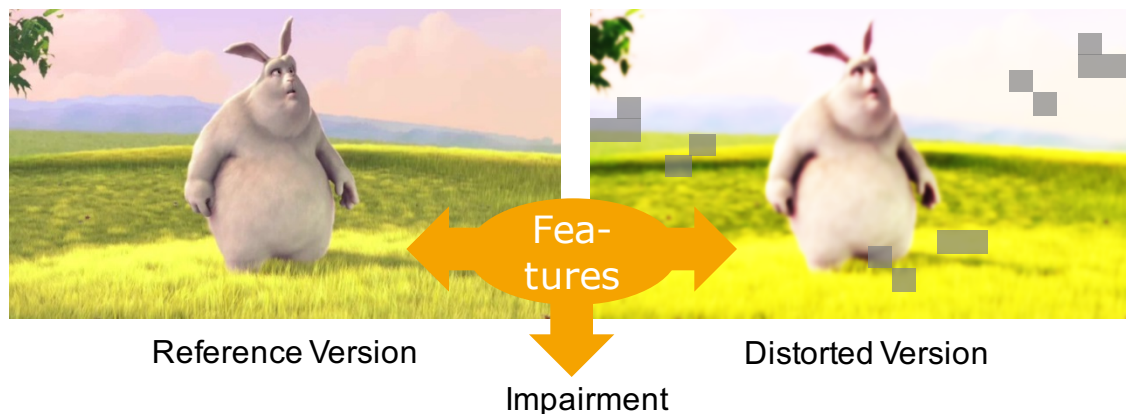
- ❖ Combinatorial explosion of different codecs and quality versions
 - YouTube: ~20 versions per video/Netflix: ~150 versions per video
- ❖ Adaptation along multiple paths (spatial/temporal resolution, quantization)
 - Perceived quality for different adaptation paths is related differently

Goal: Precise metrics for guiding video adaptation algorithms in real-time.

1. Background: Full-Reference Metrics

❖ How to define “quality” in video streaming?

- Full reference: quality difference between distorted and undistorted video based on visual characteristics



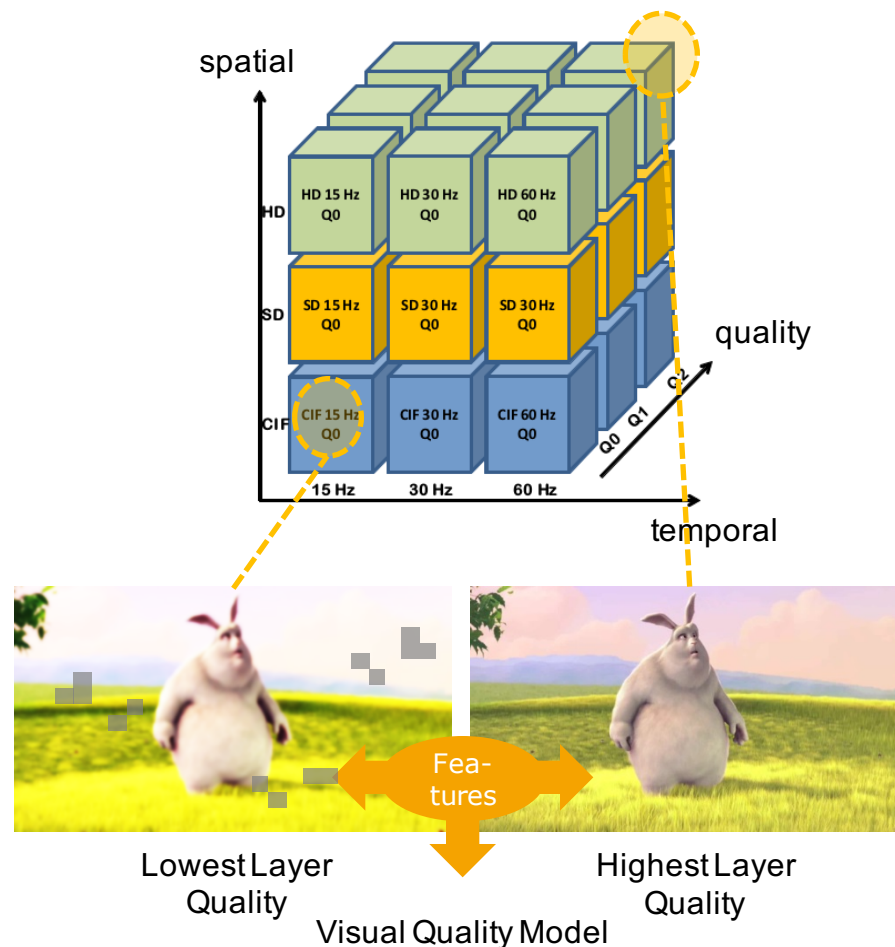
Pros: working on decoded stream, thus highly precise and (mostly) independent of encoding.

Problem: feature extraction is computationally intense, productive or real-time application (live streaming) seems out of reach.

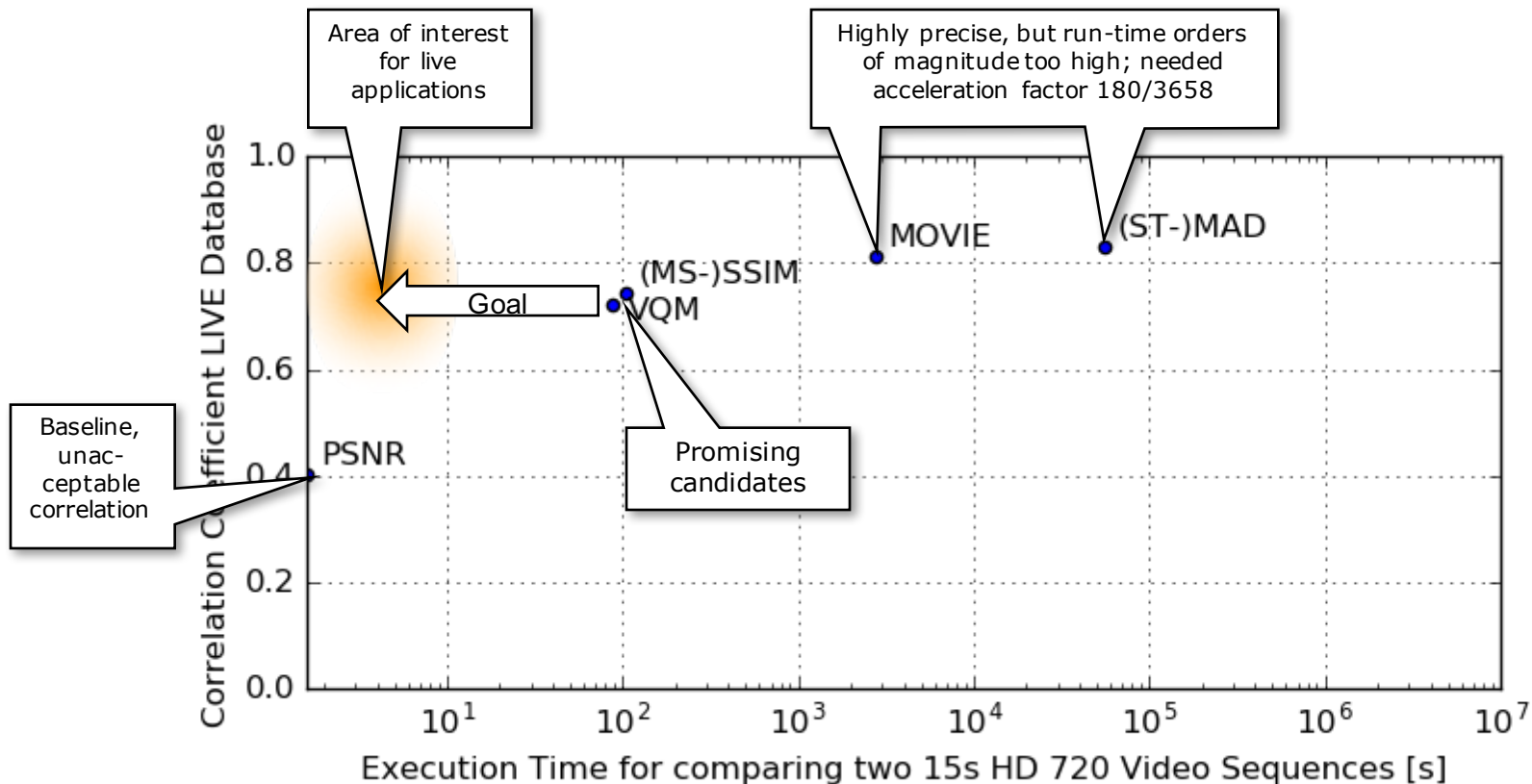
1. Background: Server Side Visual Quality Measurements

Measuring on top of an adaptive stream

- ❖ Highest layer sequence serves as reference version
- ❖ Impairment value for all other layer is calculated using Full Reference metric and sent along with the stream
 - e.g. in manifest file
- ❖ Client adapts according to annotated visual quality

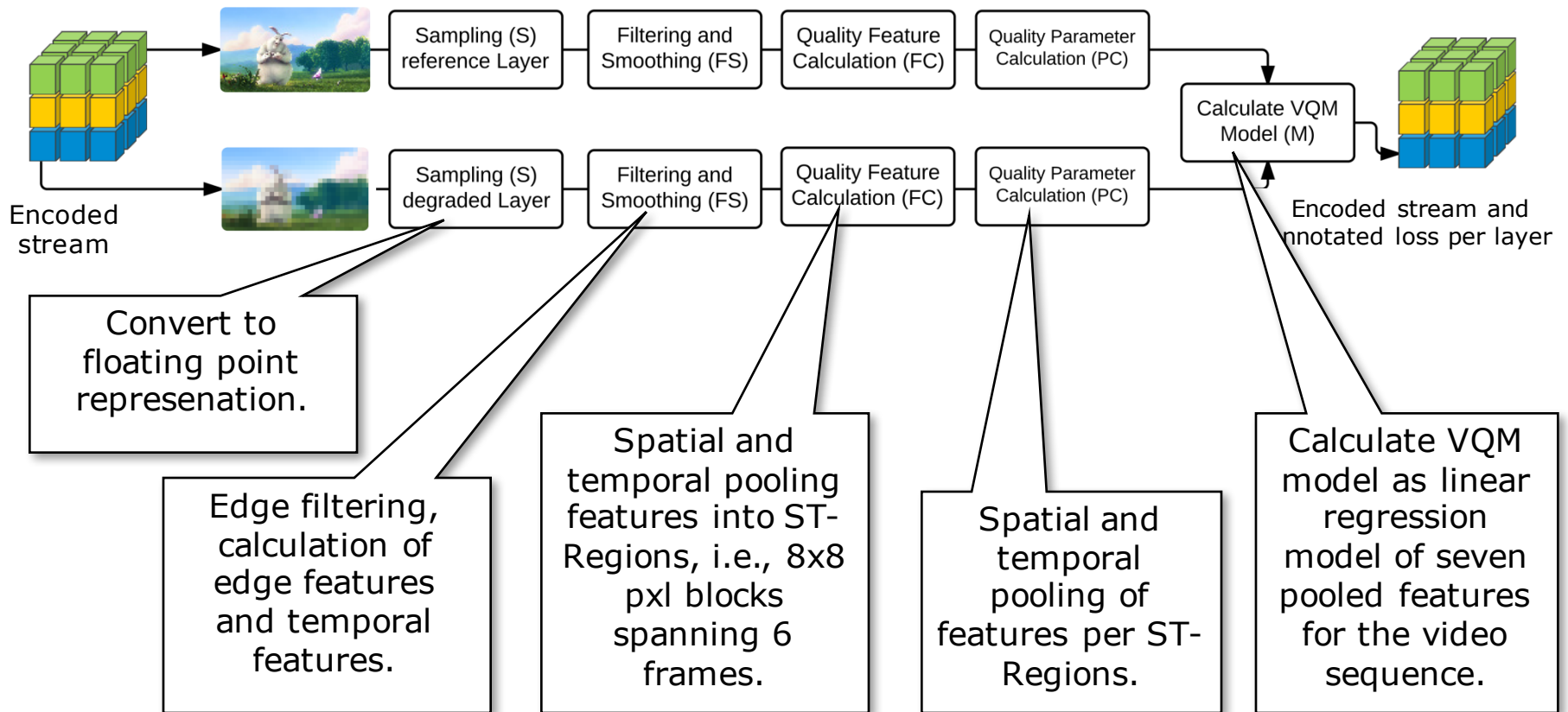


2. Related Work: Existing Full Reference (FR) Metrics



VQM is chosen as a base for this work, as it is widely accepted and standardized by the NTIA.

3. VQM Analysis



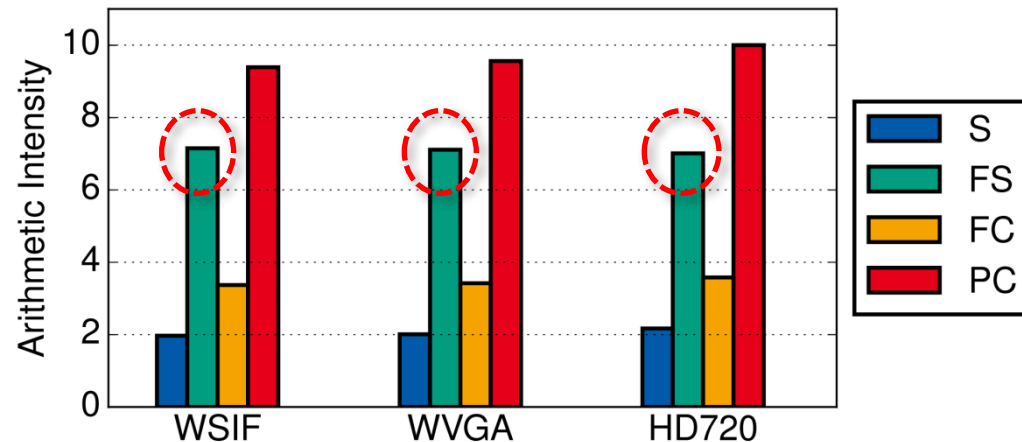
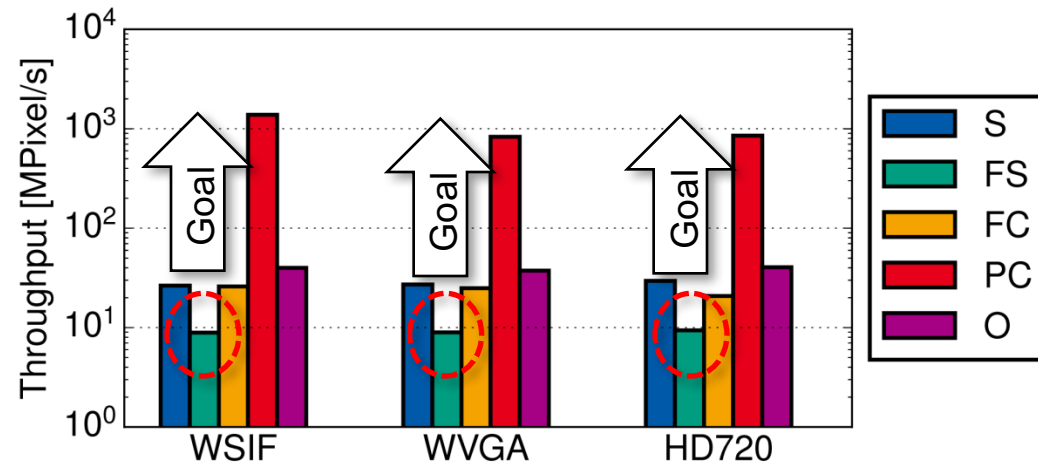
4. VQM Analysis

Step-wise VQM Performance Analysis on CPU

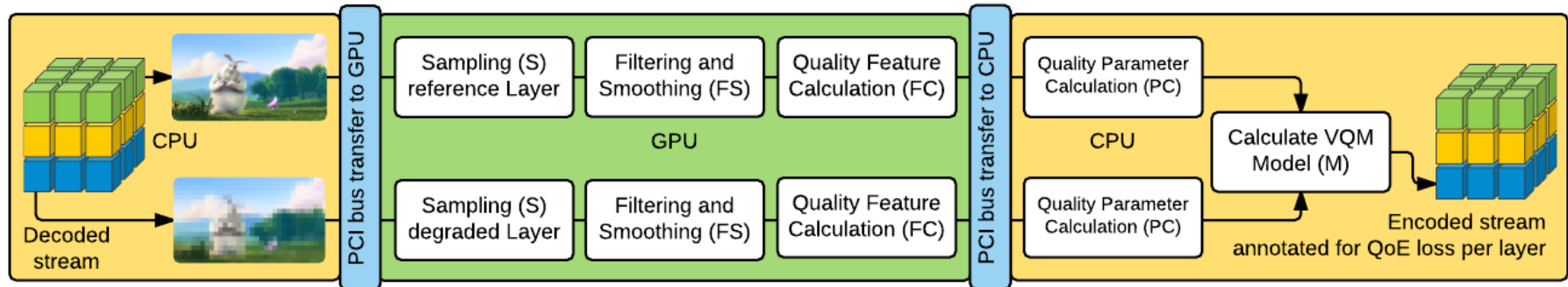
- ❖ Throughput in MPixel/s
 - Higher is better
- ❖ Arithmetic Intensity
 - Ratio of Floating Point Operations and memory accesses

Promising steps for GPU implementation

- ❖ Low Throughput and high Arithmetic Intensity (especially **FS**)
- ❖ Identifies steps bound by computational performance, not by I/O performance
- ❖ **Goal:** remove performance bottlenecks of **S-FS-FC** steps



5. Parallel Design



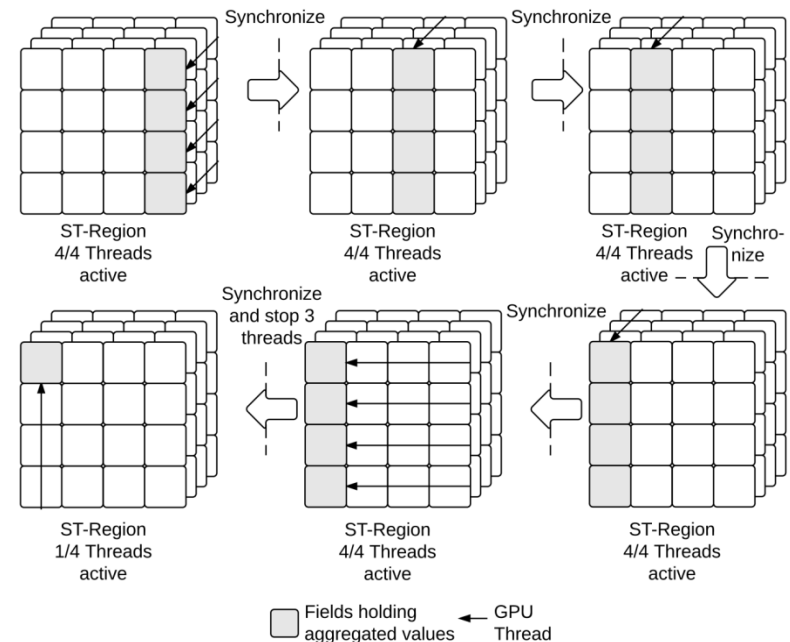
S-FS-FC steps are moved to GPU

- ❖ Heavy refactoring and parallel optimization to fully utilize GPU architecture
 - **Optimized parallel pooling of features**
 - Memory access optimization
 - Texture caching and hardware interpolation of values where needed
 - Utilization of on-chip memory where possible to prevent RAM access
 - Increases time spent on number crunching vs IO operations
 - **Decomposition of algorithms into smaller units**
- ❖ Adding support for spatial and temporal downsampled video

6. Optimizations (excerpt)

VQM pools (aggregates) features several times

- **Problem:** GPU executes batches of threads; what is the optimal number of threads?
- Optimized feature pooling:
 - **87.5% hardware utilization vs. 44% hardware utilization in worst case**



Decomposition of convolutions to multiple GPU kernel functions

- **Problem:** doing full convolutions degrades hardware utilization
 - High number of threads consuming registers
- Decomposed convolutions bring considerable performance benefits

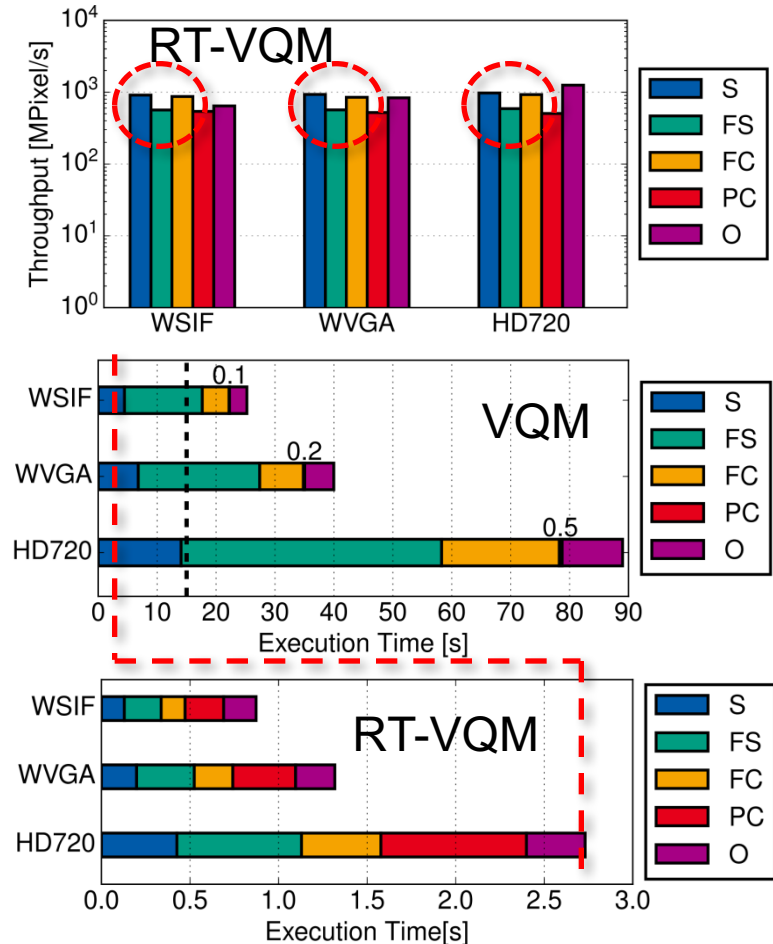
7. Evaluation

Single precision or double precision?

- ❖ GPU handles single precision much faster
- ❖ RT-VQM in single precision mode vs. VQM in double precision produces negligible error
 - Total error $< 1 \times 10^{-5}$

Throughput and execution time

- ❖ All steps of the pipeline can process 10^3 MPixel/s
- ❖ RT-VQM beats VQM by a factor of ~ 27 -30 by runtime ...
- ❖ ... and by a factor of ~ 7 -8 normalized by cores

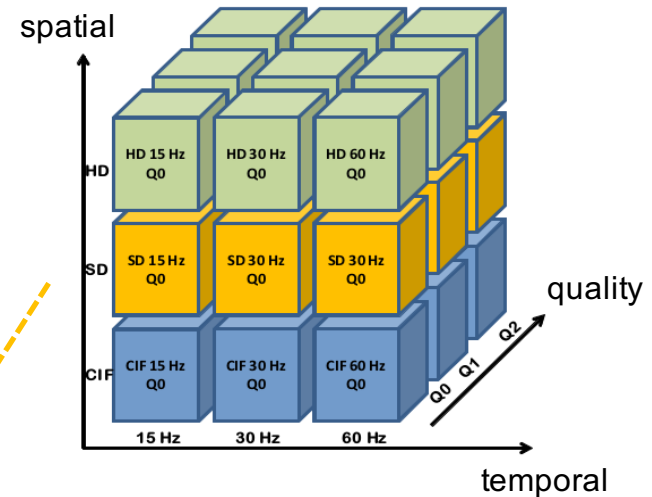


8. Evaluation

RT-VQM runtime can be reliably predicted

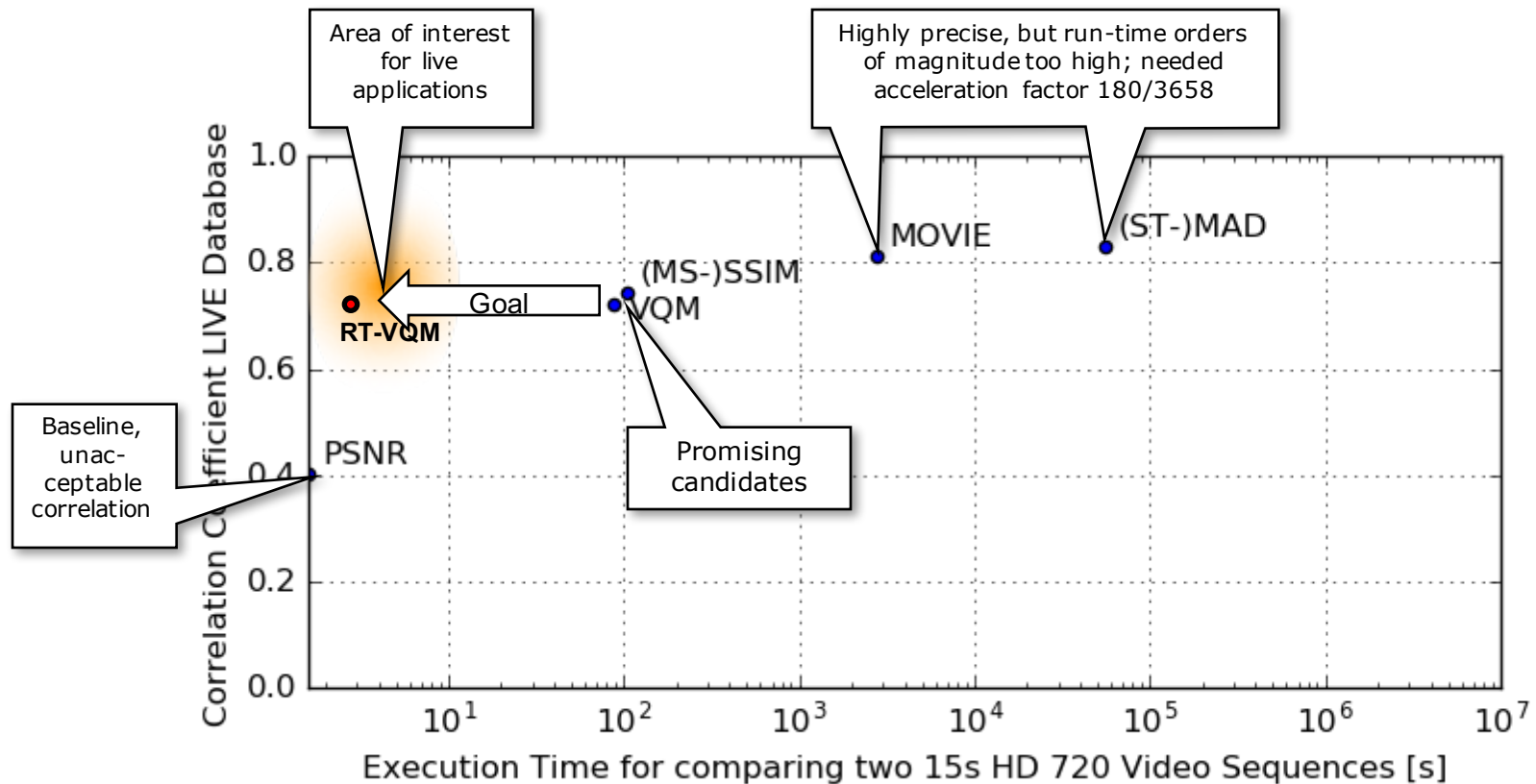
- ❖ Depends on resolution and framerate of reference and number of representations only

	WSIF	WVGA	HD720
Single Precision	30	20	9
Double Precision	24	16	7



- ❖ Up to nine representations in real-time using single precision
- ❖ Measured on GeForce GTX 780 Ti, code should scale well with new generations of hardware

9. Conclusions



10. Outlook

Visual quality guided adaptation of adaptive live video streams

Visual quality aware caching

- Calculate impairment for blocks being transferred through the cache, deliver different same quality segments if present in cache

Streaming policies

- E.g., service differentiation with guaranteed visual lower service boundaries

11. Discussion

- ❖ Thank you for your attention
- ❖ Feel free to fork the code on github for your own purposes ...
 - <https://github.com/mwichtlh/rt-vqm>

