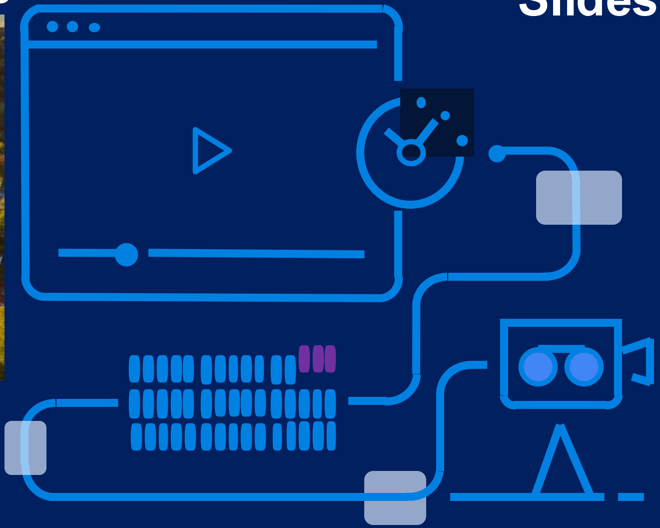
UNIVERSITÄT KLAGENFURT

ATHENA
Christian Doppler (CD) Laboratory
https://athena.itec.aau.at/

Slides

Video Coding Advancements in HTTP Adaptive Streaming



IEEE ICME 2025, June 30, 2025

Hadi Amirpour
Christian Timmerer

Alpen-Adria Universität Klagenfurt, Austria
Christian Doppler Laboratory ATHENA

Presenter Today

Hadi Amirpour

- Focus and Research Area
  - Video Streaming
  - Video Compression
  - Quality of Experience

Asst.-Prof. at Alpen-Adria-Universität Klagenfurt
Web: https://hadiamirpour.github.io/

- Education
  - Ph.D. (Dr.-techn.) in Computer Science – Universität Klagenfurt
  - M.Sc. in Electrical and Electronics Engineering – K. N. Toosi University
  - B.Sc. in Biomedical Engineering – Islamic Azad University
  - B.Sc. in Electrical Engineering – Amirkabir University of Technology

Presenter Today

Christian Timmerer

- Focus and Research Area
  - Multimedia Systems
  - Adaptive Video Streaming (ABR)
  - Quality of Experience

- Education and Experiences

  2003: MSc CS (Dipl.-Ing.)
  2006: PhD CS (Dr.-techn.)
  2012: Co-founded Bitmovin
  2014: Habilitation (Priv.-Doz.) & Assoc. Prof.
  2016: Dep. Director @ ITEC/AAU
  2019: Director @ ATHENA
  2022: Univ.-Prof. for Multimedia Systems
  2023: Director @ ITEC/AAU

ATHENA
Christian Doppler (CD) Laboratory

UNIVERSITÄT KLAGENFURT

Bitmovin MPEG-DASH

Univ.-Prof. at Alpen-Adria-Universität Klagenfurt
Director CD Lab ATHENA | Director Institute of Information Technology
CIO | Head of Research and Standardization at Bitmovin

Web: http://timmerer.com/

Upon Attending This Talk, You Will Know About

- Introduction to and Background of HTTP Adaptive Streaming
- Video Coding for HTTP Adaptive Streaming
- Bitrate Ladder Construction for HTTP Adaptive Streaming
- Video Coding Enhancement for Online Video Streaming
- Fast Multi-rate Encoding for HTTP Adaptive Streaming
- Edge Computing Capabilities for Video Transcoding
- Quality of Experience Parameters in Video Coding and Streaming

Introduction to HTTP Adaptive Video Streaming

Video streaming is **dominating today's Internet traffic**

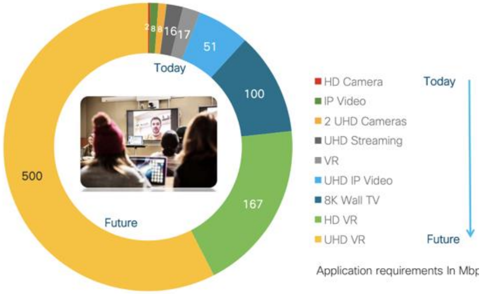
- 2024*: 68% (fixed) and 64% (mobile)
  - Video on-demand: 54% / 57% – Live: 14% / 7%
  - Main applications: YouTube, Netflix (>10%),
  - Tik Tok, Amazon Prime, Disney+ (<10%)
- Video applications are in high demand, but future ones will require even more bandwidth**
- Streaming now accounts for a larger share of total TV viewing than both broadcast and cable combined.***

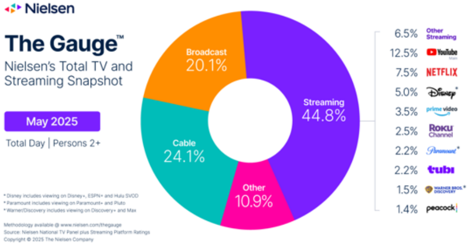
Top Content Categories by Downstream Volume – Fixed

	Downstream Volume		
	Content Category	% DS Vol	Sub. Volume
1	On-Demand Streaming	54%	7.9 GB
2	Live Streaming	14%	2.0 GB
3	File Delivery	13%	2.0 GB
4	Browsing	3%	441 MB
5	Game Play	3%	398 MB
6	Video Call	2%	300 MB
7	Messaging	0.6%	81 MB
8	Voice Call	0.5%	74 MB
9	Machine to Machine	0.01%	2 MB
10	AR/VR	0.00%	18 KB
11	Other	10%	1.4 GB

Top Content Categories by Downstream Volume – Mobile

	Downstream Volume		
	Content Category	% DS Vol	Sub. Volume
1	On-Demand Streaming	57%	900 MB
2	File Delivery	11%	173 MB
3	Live Streaming	7%	107 MB
4	Game Play	5%	75 MB
5	Video Call	5%	73 MB
6	Browsing	2%	38 MB
7	Messaging	2%	29 MB
8	Voice Call	1%	19 MB
9	Machine to Machine	0.00%	68 KB
10	AR/VR	0.00%	1 KB
11	Other	10%	160 MB





Sources:
* Sandvine Global Internet Phenomena (January 2024)
** Cisco Annual Internet Report (2018–2023) White Paper (March 2020)
*** Nielsen's The Gauge™ (May 2025)

Evolution of Video

THE PAST:

Invention of camera, still image photography, color reproduction, film, moving pictures

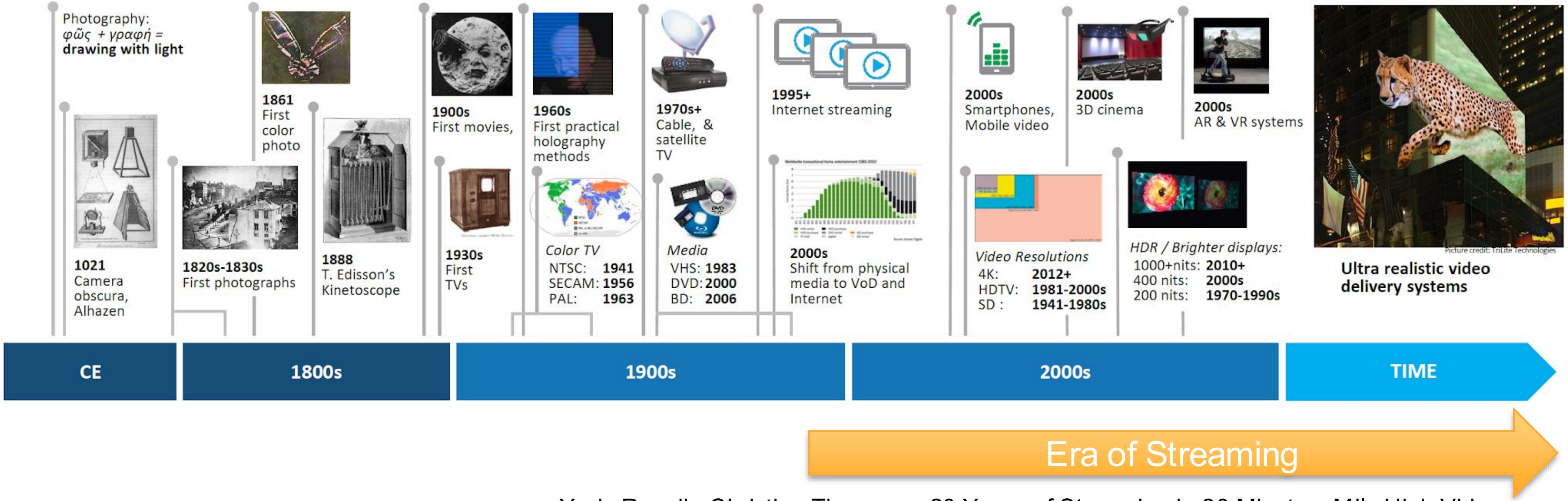
THE PRESENT:

New delivery methods: TV, recordable media, digital compressed formats, Internet streaming, mobile.

Increasing degree of realism: immersive video, 3D (holography, stereoscopic rendering, etc.)

THE FUTURE:

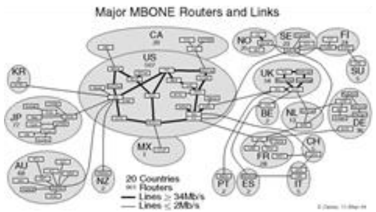
Recording & reproduction systems making rendered video undistinguishable from reality.



Yuriy Reznik, Christian Timmerer, 20 Years of Streaming in 20 Minutes, Mile-High Video 2020, https://athena.itec.aau.at/2020/11/20-years-of-streaming-in-20-minutes/

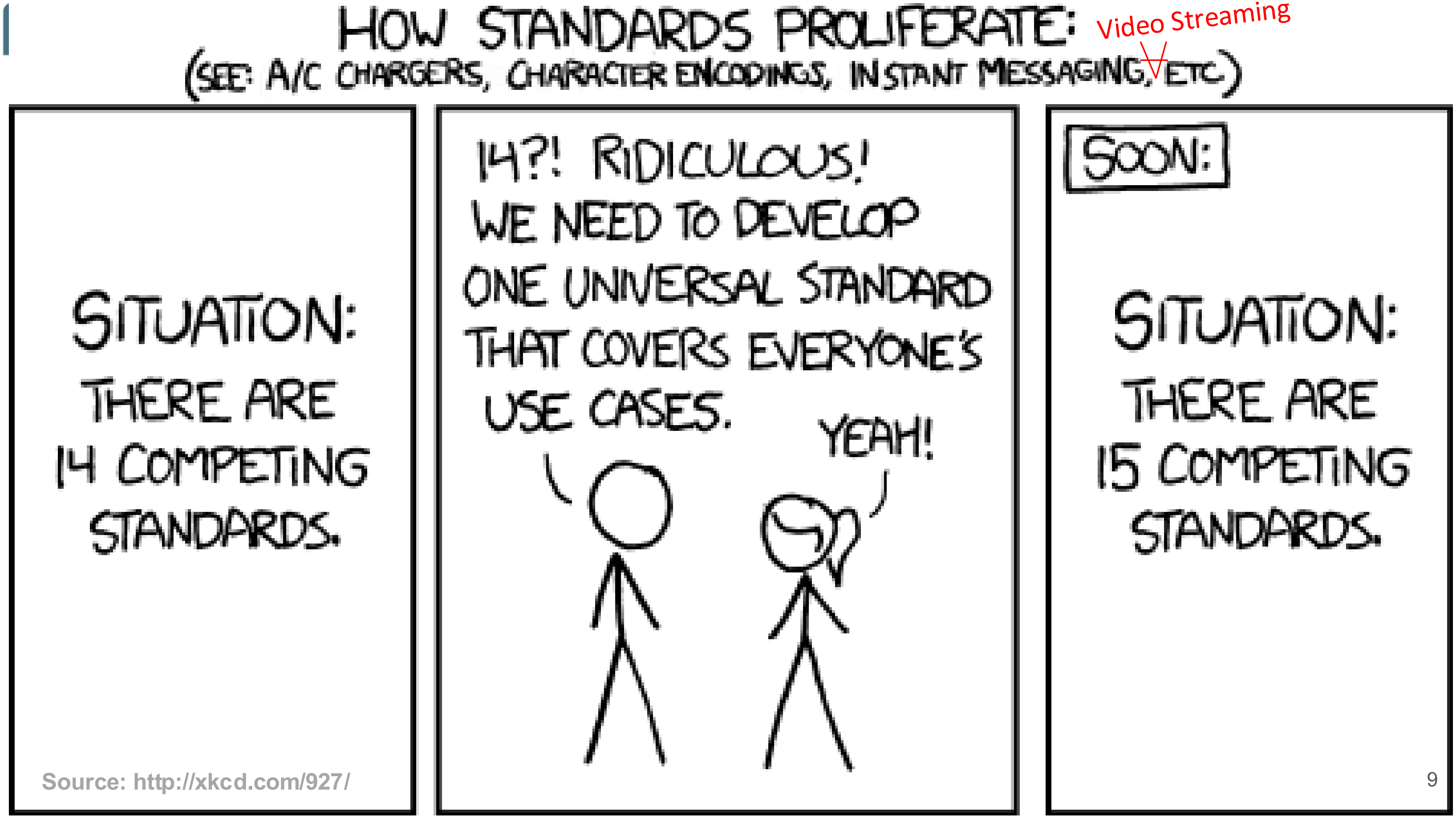
Early Streaming Systems

- **1993: Multicast Backbone (MBONE)**
  - Virtual multicast network connecting several universities & ISPs
  - RTP-based video conferencing tool (vic) is used to stream videos
  - 1994 Rolling Stones concert – first major event streamed online
- **1995: RealAudio, 1997: RealVideo**
  - First commercially successful mass-scale streaming system
  - Proprietary protocols, codecs: PNA, RealAudio, RealVideo
  - Worked over UDP, TCP, and HTTP ("cloaking" mode)
  - First major broadcast: 1995 Seattle Mariners vs New York Yankees
- **1996: VDOnet, Vivo, NetShow, VXtreme, ...**
  - Many vendors have competed in streaming space initially
  - Vivo & Xing have been acquired by Real, VXtreme by Microsoft
  - By 1998, 3 main vendors remained: Real, Microsoft , and Apple
- **1998: RealSystem G2**
  - First **Adaptive BitRate (ABR)** streaming system



“The nice thing about standards is that you have so many to choose from.”

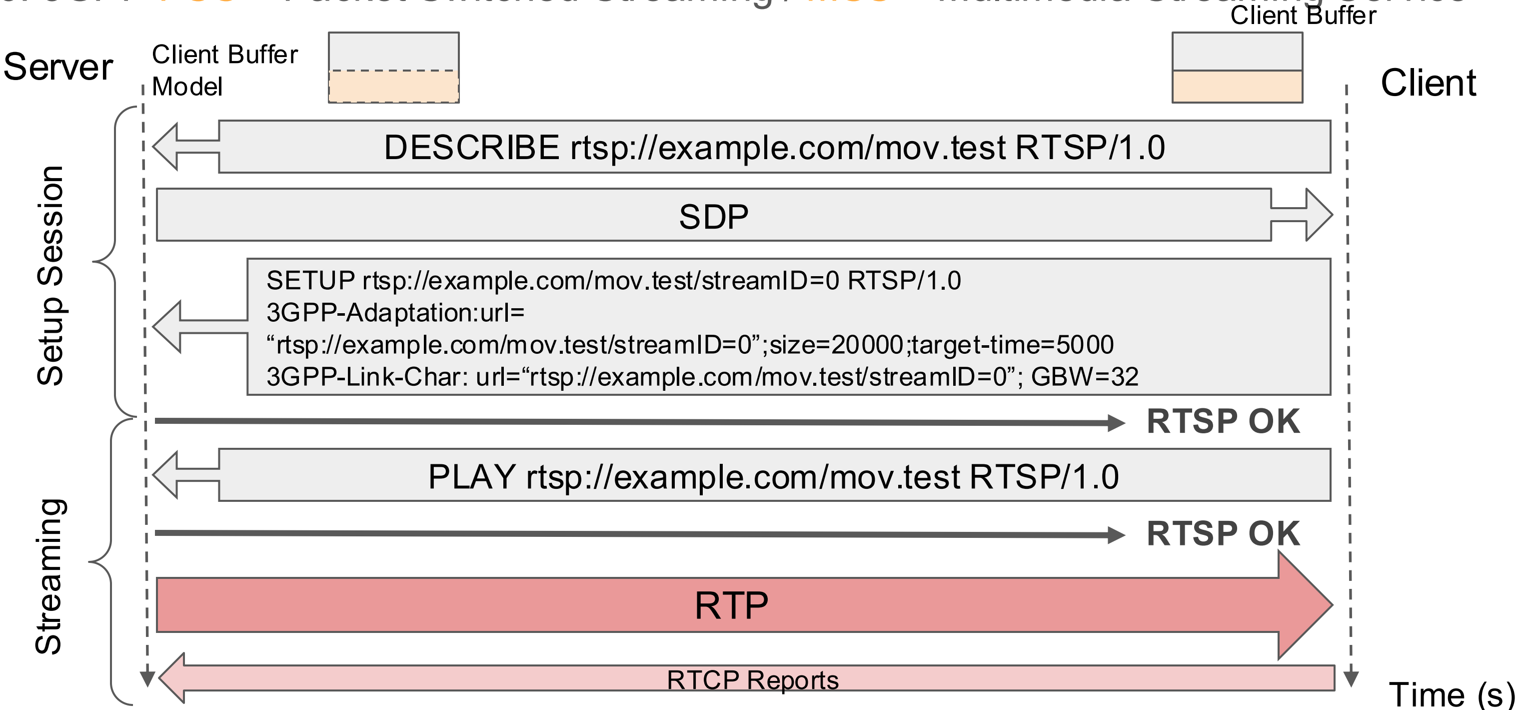
Andrew S. Tanenbaum, Computer Networks



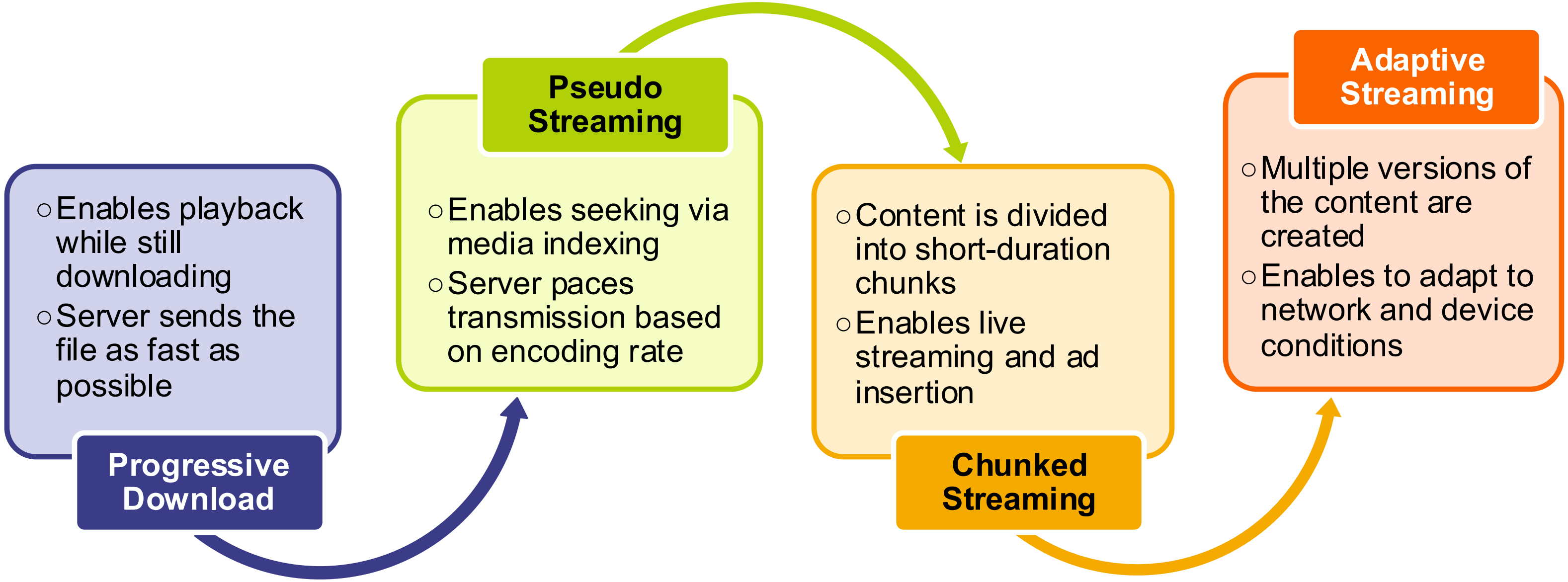
Source: http://xkcd.com/927/

Early Streaming Standards

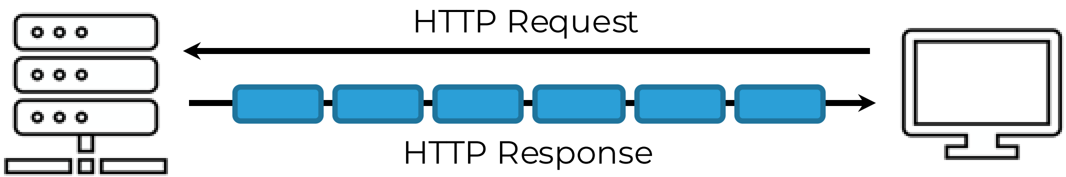
- 1996: Real Time Streaming Protocol (RTSP) et al.
- 2000: ISMA – Internet Streaming Media Alliance
- 2006: 3GPP PSS – Packet Switched Streaming / MSS – Multimedia Streaming Service

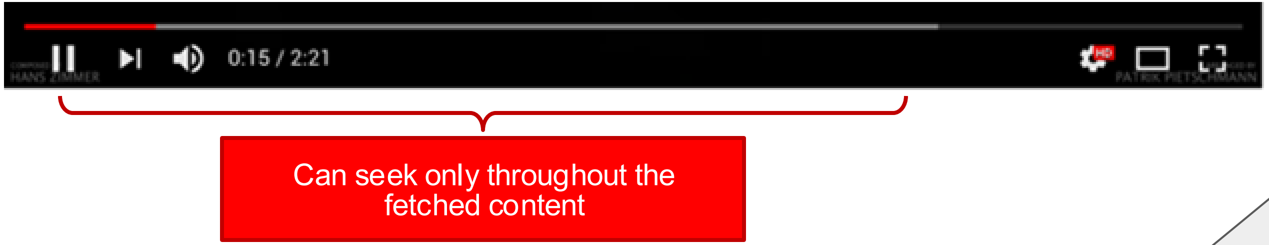


2005+: Video Delivery over HTTP/TCP



Progressive Download





"TCP generally provides good streaming performance when the achievable TCP throughput is roughly twice the media bitrate, with only a few seconds of startup delay"

Bing Wang, Jim Kurose, Prashant Shenoy, and Don Towsley. 2004. **Multimedia streaming via TCP: an analytic performance study**. ACM International Conference on Multimedia (MM'04). DOI:https://doi.org/10.1145/1027527.1027735

Streaming is Cooler, more Viewer Friendly

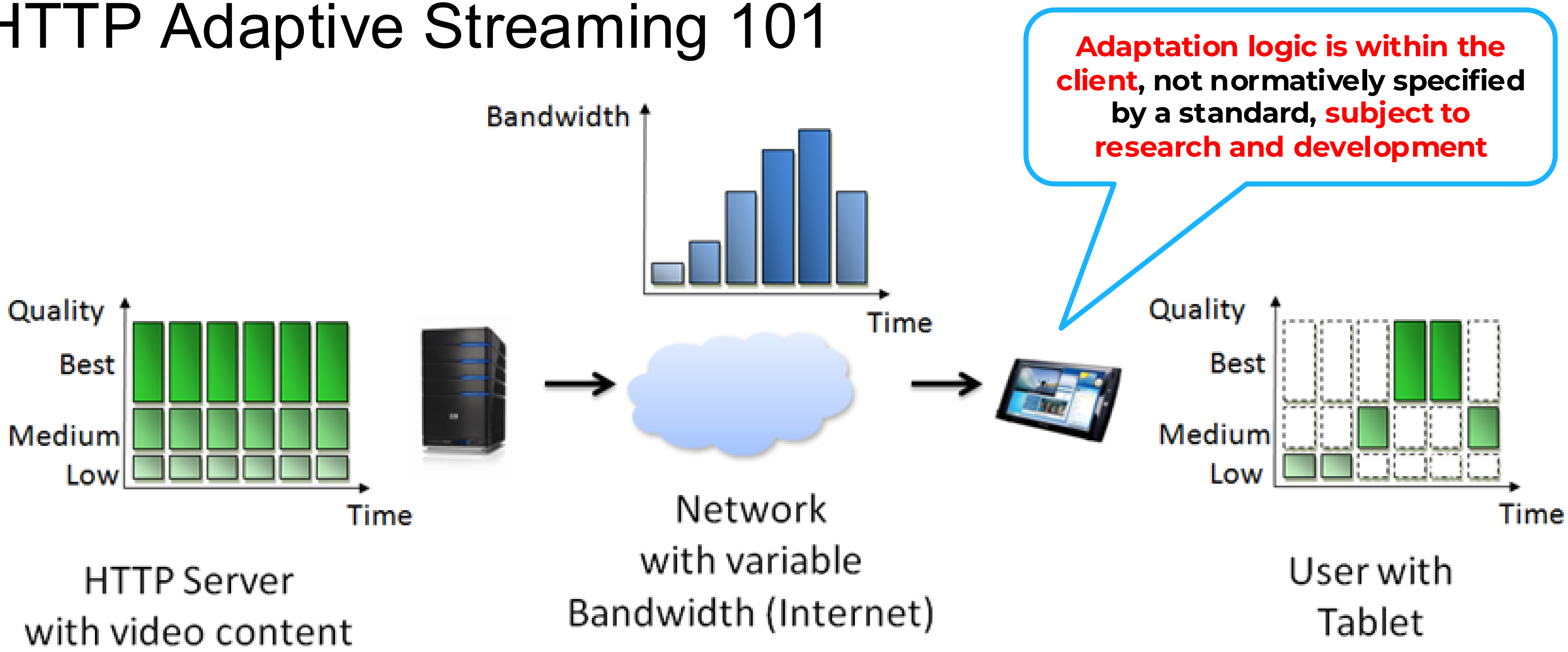
- Playback starts when there are just few seconds of data
- Download rate will match the encoding bitrate and downloading pauses if the player pauses → less waste

Can seek to anywhere in the entire content

Implements "pseudo streaming" and mimics RTSP-style streaming, but via HTTP

L. De Cicco, S. Mascolo. **An Experimental Investigation of the Akamai Adaptive Video Streaming.** *in Proc of USAB 2010, special session Interactive Multimedia Applications (WIMA)*, Klagenfurt, Austria, 3-4 November 2010, LCNS 6389, pp. 447-464, Springer-Verlag, doi:10.1007/978-3-642-16607-5

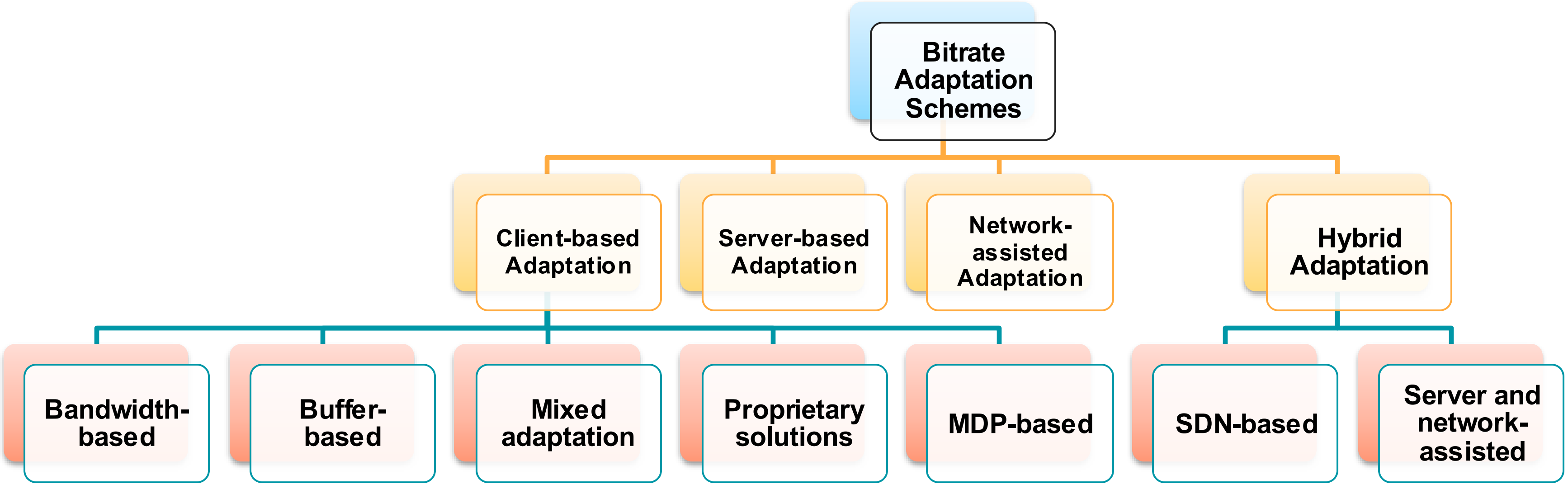
HTTP Adaptive Streaming 101



Christian Timmerer and Hermann Hellwagner. 2020. **HTTP Adaptive Streaming – Where Is It Heading?**. In Brazilian Symposium on Multimedia and the Web (WebMedia '20), November 30-December 4, 2020, São Luís, Brazil. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3428658.3434574

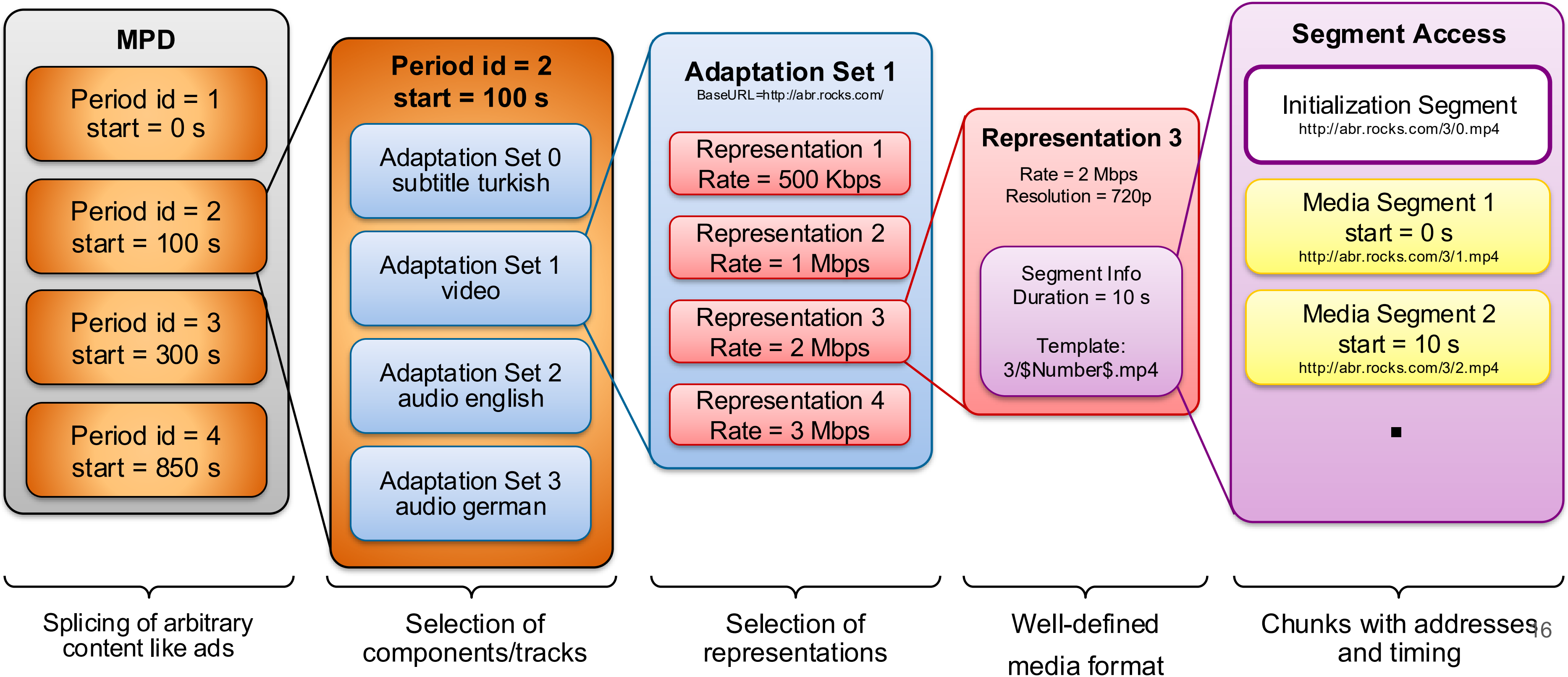
Christian Timmerer, Hadi Amirpour, Farzad Tashtarian, Samira Afzal, Amr Rizk, Michael Zink, and Hermann Hellwagner. 2025. **HTTP Adaptive Streaming: A Review on Current Advances and Future Challenges**. ACM Trans. Multimedia Comput. Commun. Appl. Just Accepted (May 2025). https://doi.org/10.1145/3736306

Bitrate Adaptation Schemes



A. Bentaleb, B. Taani, A. C. Begen, C. Timmerer and R. Zimmermann, "**A Survey on Bitrate Adaptation Schemes for Streaming Media Over HTTP,**" in IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 562-585, Firstquarter 2019. doi: 10.1109/COMST.2018.2862938

MPEG-DASH Data Model



UNIVERSITÄT KLAGENFURT

MPEG-DASH Status (11/2024)

MPEG-DASH 1st Edition 23009-1:2012

- Events
- Asset Identifier

2nd Edition 23009-1:2014 ✓

AMD1
- Server-client NTP sync
- Extended profiles ✓

AMD2
- SRD
- URL param insertion
- Role extensions ✓

AMD3
- AuthN/AuthZ
- NTP anchor
- External MPD link
- Period continuity
- Generalized HTTP header extensions & queries ✓

3rd Edition 23009-1:2019

AMD4
- Flexible segment & Broadcast TV profile
- MPD chaining
- MPD fallback
- Preselections
- Data URLs in MPD
- Labels ✓
- Switching x adaptation sets

AMD5 (AMD1 to 3rd edition)
Device information, quality equivalence descriptor, timed text roles, announcing popular content, flexible IOP signaling, early available periods, signaling missing/alternative segments

4th Edition 23009-1:2020 ✓

AMD1 (to 4th edition)
CMAF support, event/timed metadata processing, resynchronization, patch method for MPD updates, preroll

5th Edition 23009-1:2022 ✓

NEW!

AMD1+2+3 (to 5th ed)
Preroll, nonlinear playback, EDRAP, seg. seq. for rand. access and switching, CMCD, content steering

6th Edition 23009-1:20xx

23009-2 Conformance and Reference Software

libdash update to 5th edition

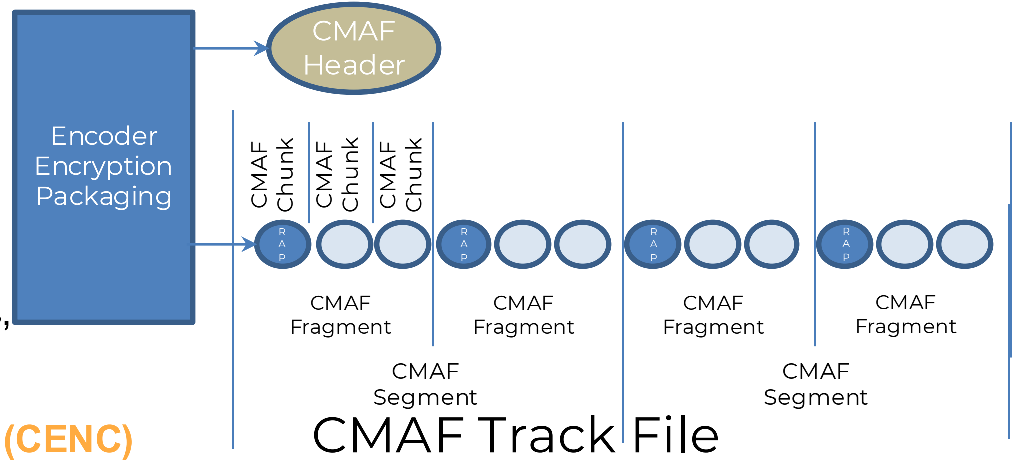
3rd Edition

- 23009-4 Segment Encryption & Authentication ✓
- 23009-5 Server & Network Assisted DASH ✓
- 23009-6 Full Duplex DASH ✓
- 23009-8 Session based DASH operations ✓
- 23009-9 Redundant Encoding and Packaging (REAP) ✓

NEW!

Additional Tools under development
- Alternative MPD events
- CMCD integration
- MV-HEVC support

Common Media Application Format (CMAF)



Media delivery has **three main** components:

- Media format
- Manifest
- Delivery

CMAF defines the **media format only** (fragments, headers, segments, chunks, tracks)

CMAF uses **ISOBMFF** and **common encryption (CENC)**

- CENC means the media fragments can be decrypted/decoded using different DRMs
- CMAF does not mandate CTR or CBC mode

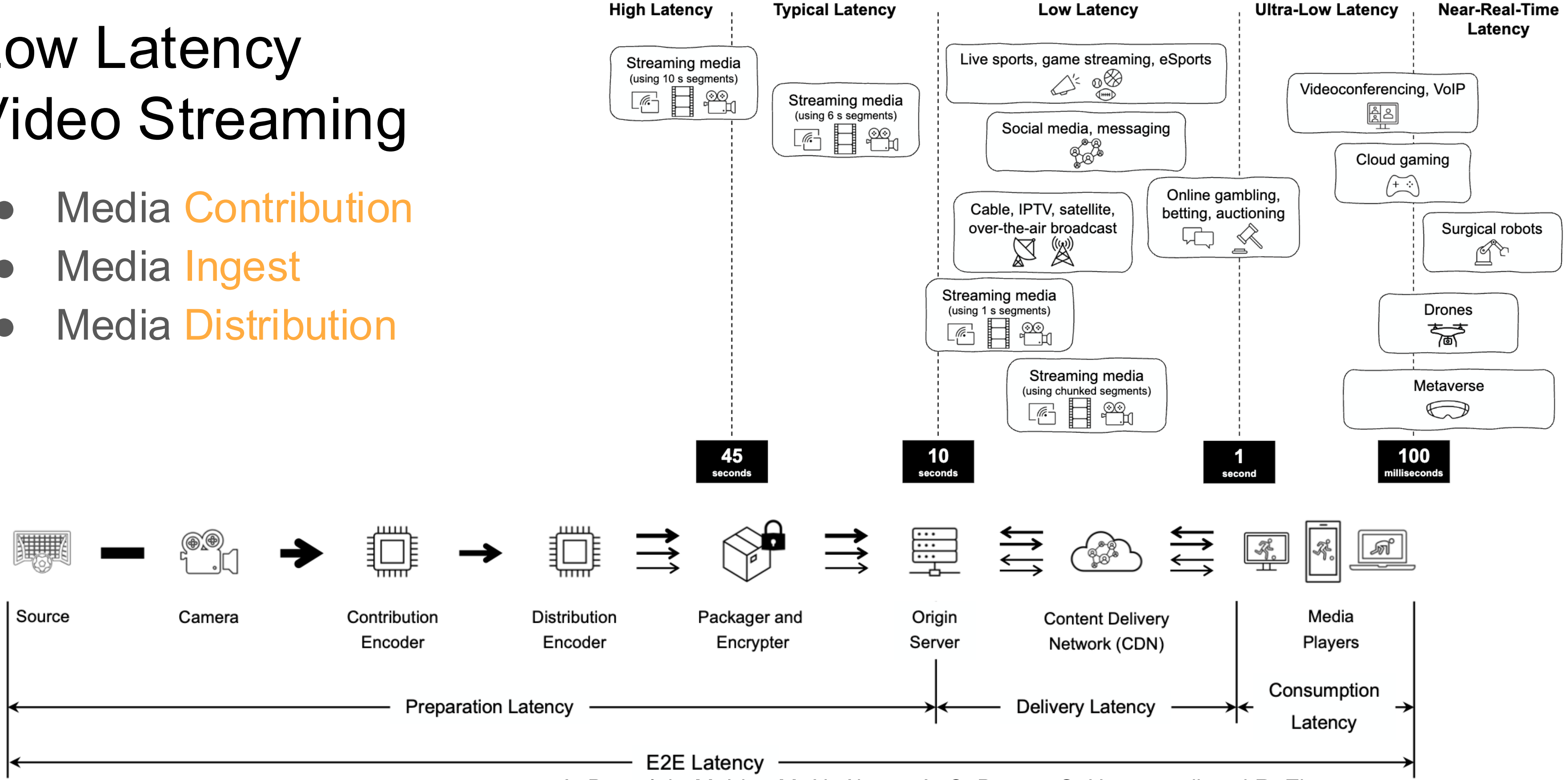
Any delivery method may be used for delivering CMAF content: HTTP, RTP multicast/unicast, LTE broadcast

CMAF is a prerequisite for **low latency HAS** (i.e., DASH-LL, LL-HLS)

Abdelhak Bentaleb, Christian Timmerer, Ali C. Begen, and Roger Zimmermann. 2020. **Performance Analysis of ACTE: A Bandwidth Prediction Method for Low-latency Chunked Streaming**. ACM Trans. Multimedia Comput. Commun. Appl. 16, 2s, Article 69 (July 2020), 24 pages. DOI:https://doi.org/10.1145/3387921

Low Latency Video Streaming

- Media Contribution
- Media Ingest
- Media Distribution



A. Bentaleb, M. Lim, M. N. Akcay, A. C. Begen, S. Hammoudi and R. Zimmermann, "**Toward One-Second Latency: Evolution of Live Media Streaming**," in IEEE Communications Surveys & Tutorials, doi: 10.1109/COMST.2025.3555514.

CMCD and CMSD in Video Streaming

What is CMCD? – Common Media Client Data

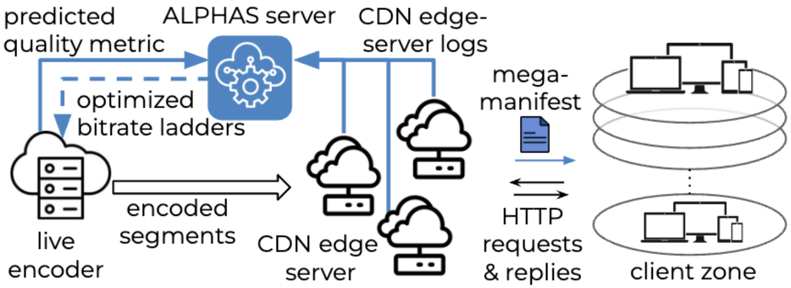
- Client ➔ Server reporting during HTTP adaptive streaming
- Sends playback metrics (buffer level, bitrate, throughput) with HTTP requests
- Used for delivery optimization, CDN prefetching, and QoE analytics
- Standard: CTA-5004

Why do they matter?

- Enable more efficient streaming over HTTP/TCP
- Support QoE improvements without intrusive client instrumentation
- Facilitate data-driven delivery optimization in modern video services

What is CMSD? – Common Media Server Data

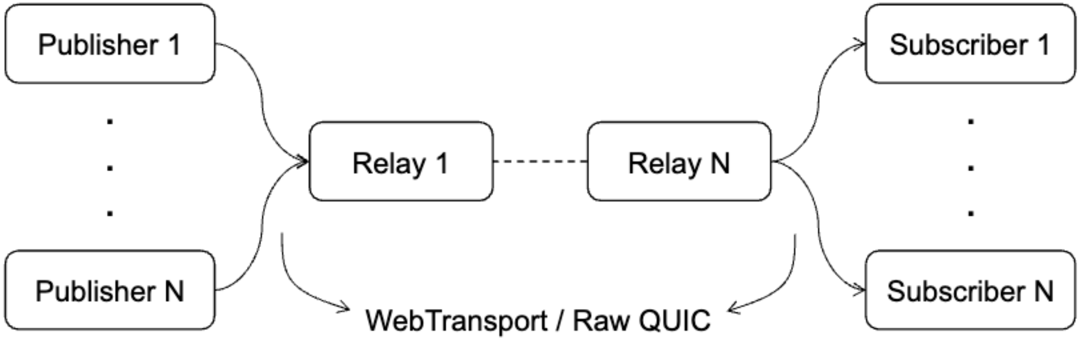
- Server ➔ Client signaling during HTTP adaptive streaming
- Provides server state (cache status, load) in HTTP responses
- Enables informed ABR (adaptive bitrate) decisions by clients
- Standard: CTA-5006



Tashtarian, Farzad, et al. "**ALPHAS: Adaptive Bitrate Ladder Optimization for Multi-Live Video Streaming.**" *IEEE International Conference on Computer Communications*, 2025.

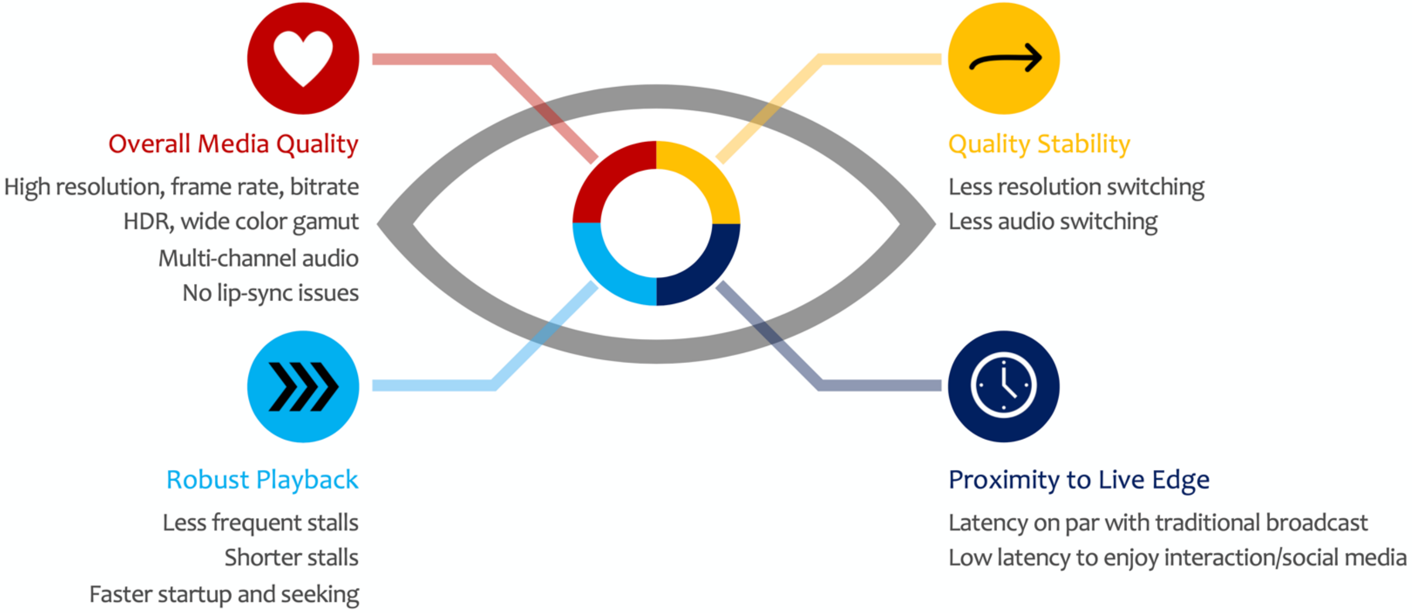
Media over QUIC

- Evolution of **HTTP/1.1** → **HTTP/2** (based on SPDY) → **HTTP/3** (QUIC)
- **Media over QUIC (MOQ)**
  - Media over QUIC Transport [draft-ietf-moq-transport-12]
  - Low Overhead Media Container [draft-ietf-moq-loc-00]
  - WARP Streaming Format [draft-ietf-moq-warp-00]
- Initial results & **testbeds** available: https://moqtail.dev/



Zafer Gurel, Tugce Erkilic Civelek, Deniz Ugur, Yigit K. Erinc, and Ali C. Begen. 2024. **Media-over-QUIC Transport vs. Low-Latency DASH: a Deathmatch Testbed**. In Proceedings of the 15th ACM Multimedia Systems Conference (MMSys '24). Association for Computing Machinery, New York, NY, USA, 448–452. https://doi.org/10.1145/3625468.3652191

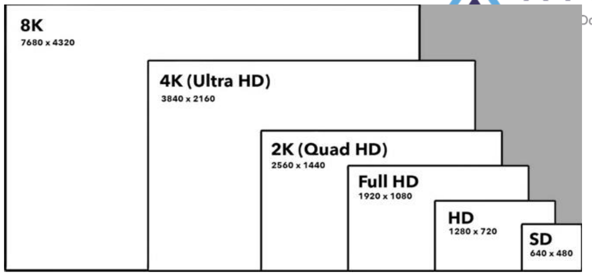
What Matters to the Human Eye?

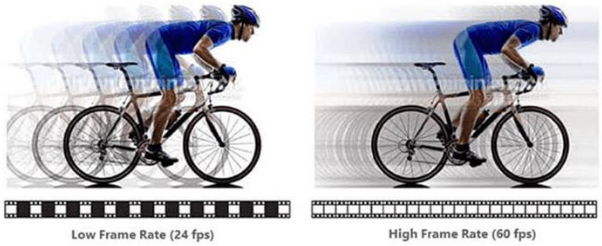


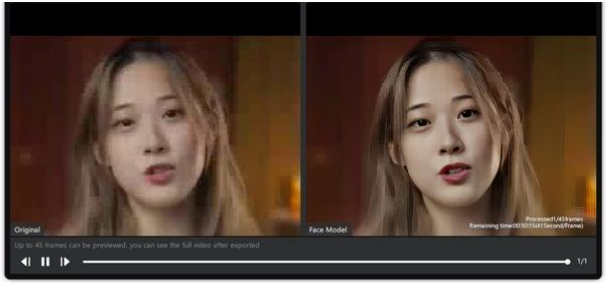
What Matters to the Human Eye?

Overall Media Quality

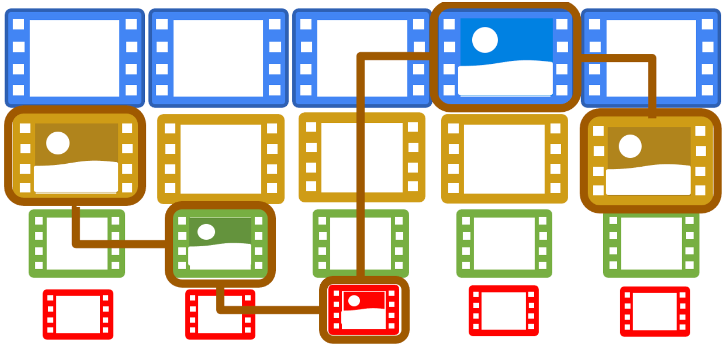
High resolution, frame rate, bitrate
HDR, wide color gamut
Multi-channel audio
No lip-sync issues







What Matters to the Human Eye?

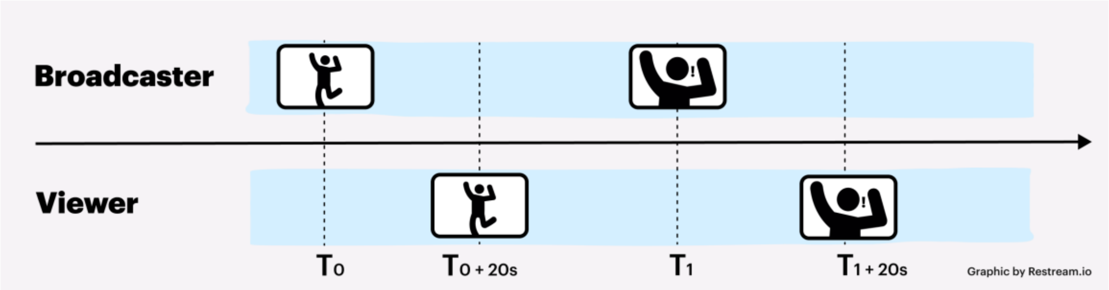


Quality Stability

Less resolution switching

Less audio switching

What Matters to the Human Eye?

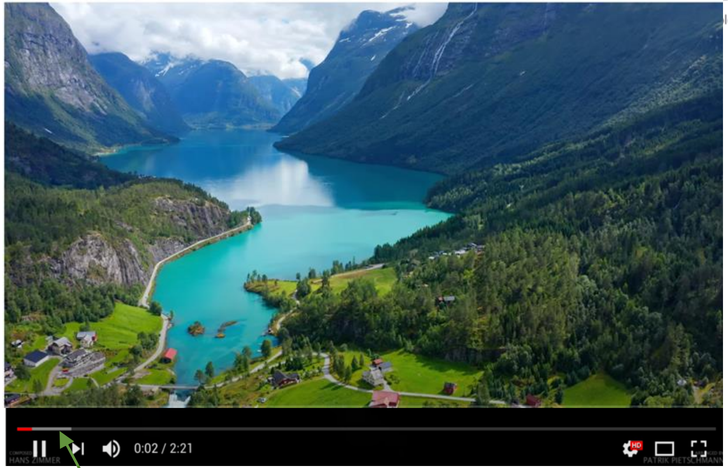


Proximity to Live Edge

Latency on par with traditional broadcast

Low latency to enjoy interaction/social media

What Matters to the Human Eye?



with video segments in buffer

without video segments in buffer

Robust Playback

Less frequent stalls

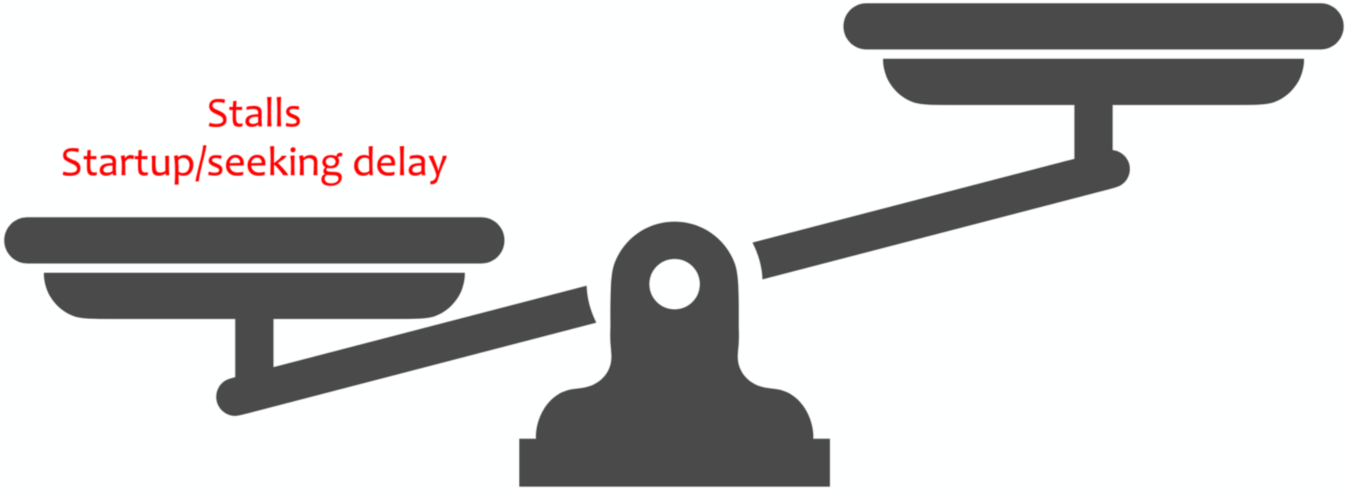
Shorter stalls

Faster startup and seeking

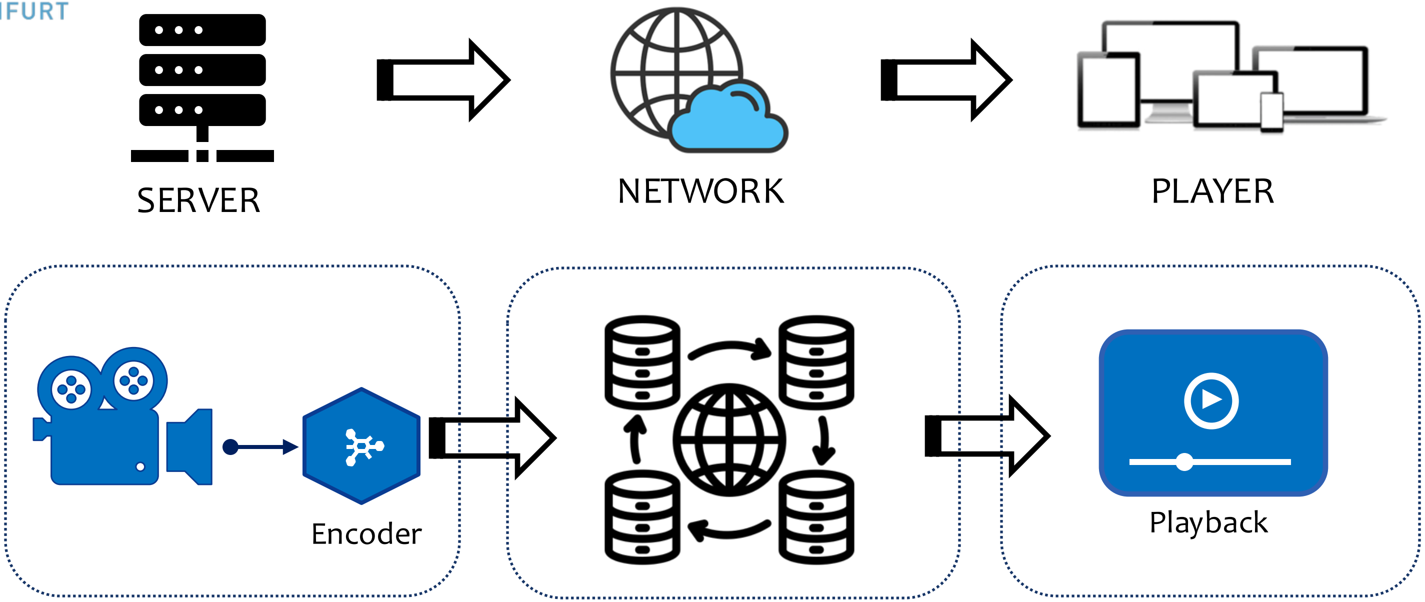
What Matters to the Human Eye?

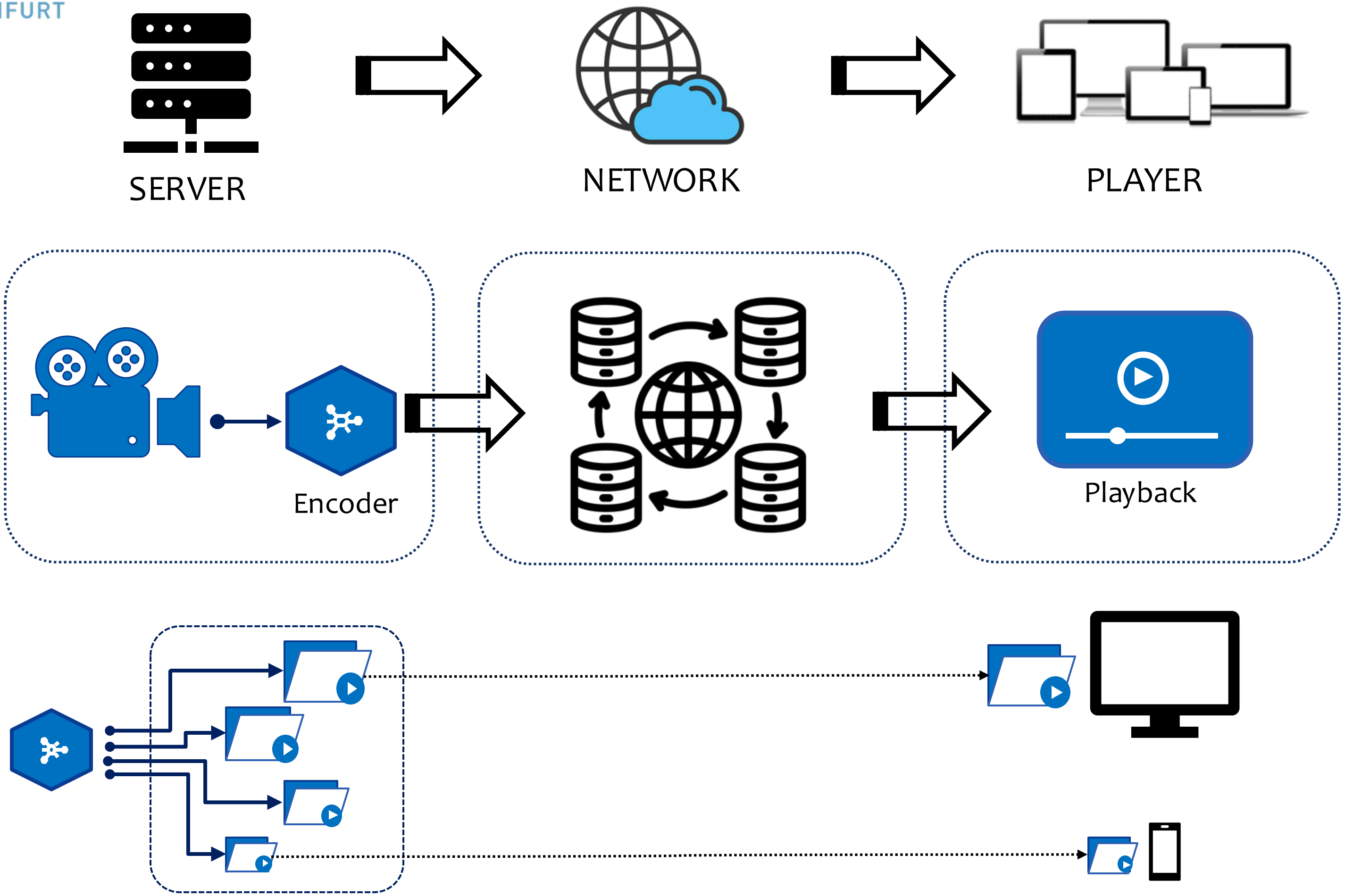
Overall quality
Quality stability
Proximity to live edge

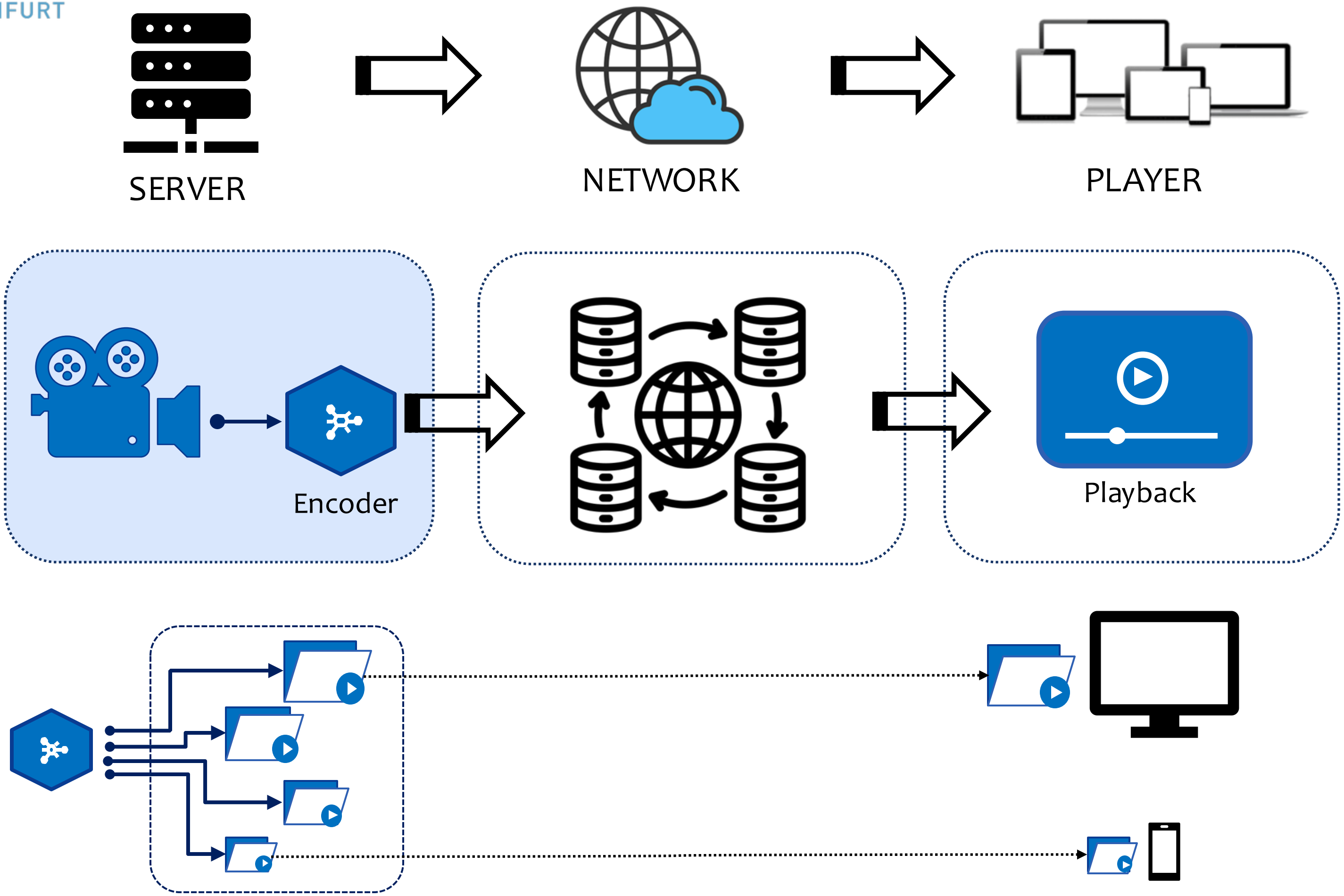
Stalls
Startup/seeking delay



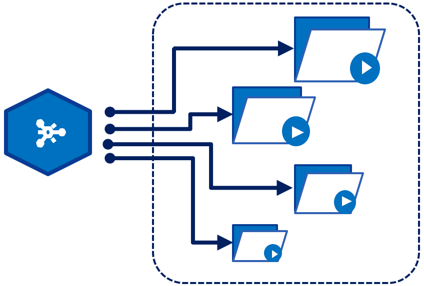

SERVER
NETWORK
PLAYER


SERVER
NETWORK
PLAYER
Encoder
Playback

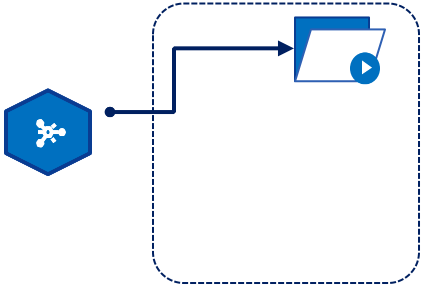

SERVER
NETWORK
PLAYER
Encoder
Playback


SERVER
NETWORK
PLAYER
Encoder
Playback

Video Coding for HTTP Adaptive Streaming (HAS)

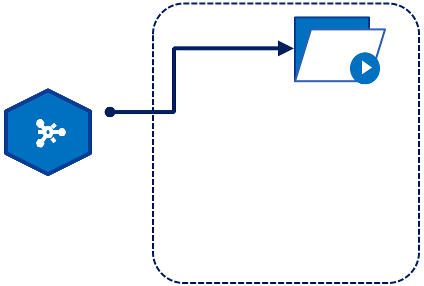


Video Coding for HTTP Adaptive Streaming (HAS)



4K, 60fps

Video Codec Timeline



AVC — 2003

The most widely adopted video codec developed by JVT: VCEG +MPEG

VP9 — 2013

A royalty-free codec developed by google and it is widely used in YouTube videos

HEVC — 2013

It achieves up to 50% bitrate savings compared to AVC developed by JVT-VC: VCEG +MPEG

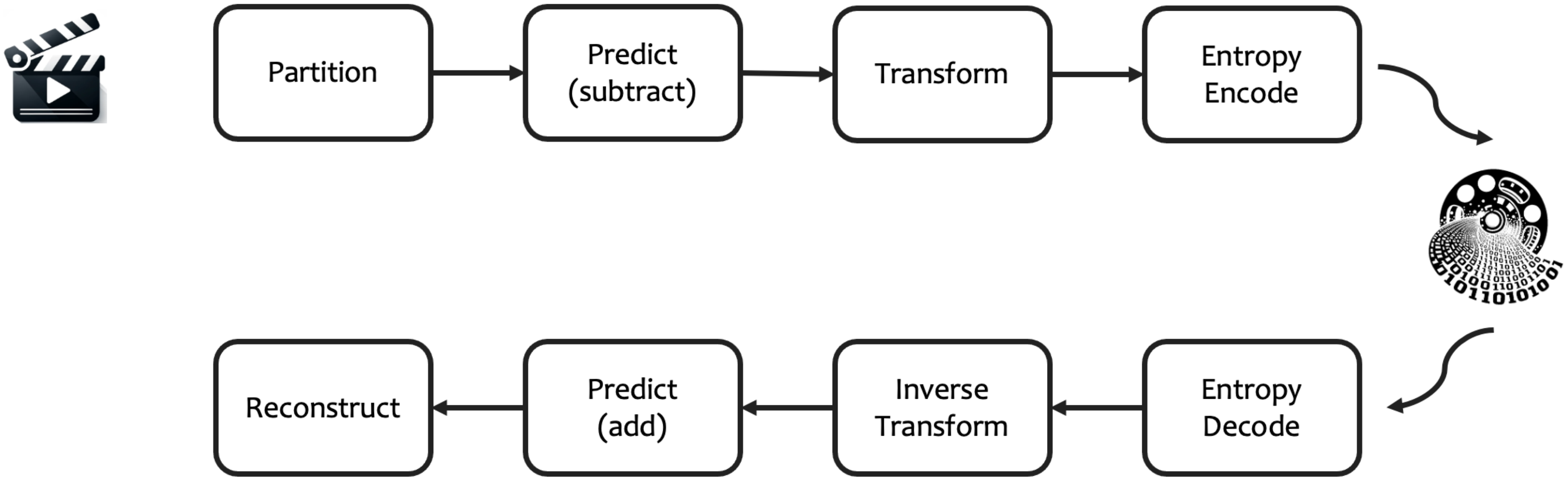
AV1 — 2018

A royalty-free codec developed by AOM as a successor to VP9, achieving more than 30% bitrate savings compared to VP9

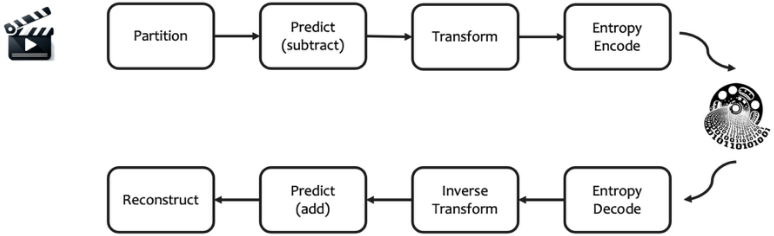
VVC — 2020

It achieve up to 50% bitrate savings compared to HEVC, known for its versatility. It was developed by JVET: VCEG+MPEG

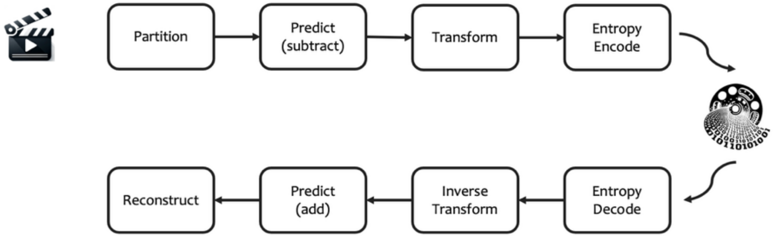
Video Coding Building Blocks

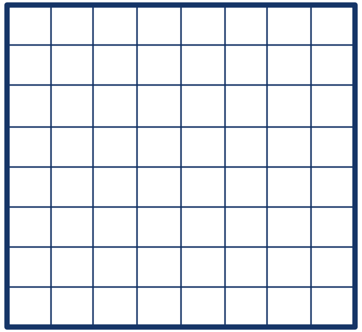


Video Coding Building Blocks

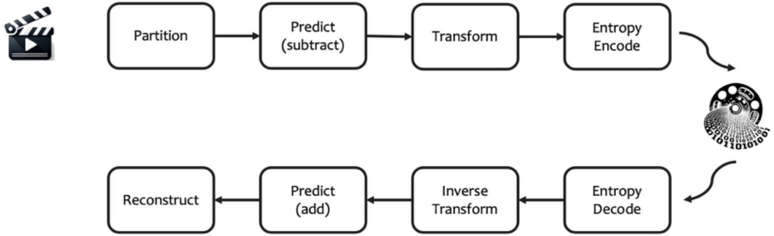


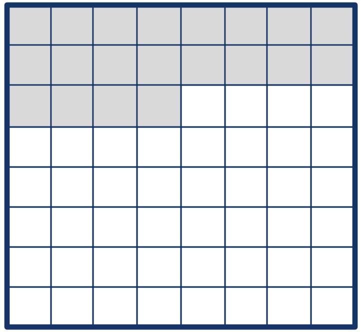
Video Coding Building Blocks



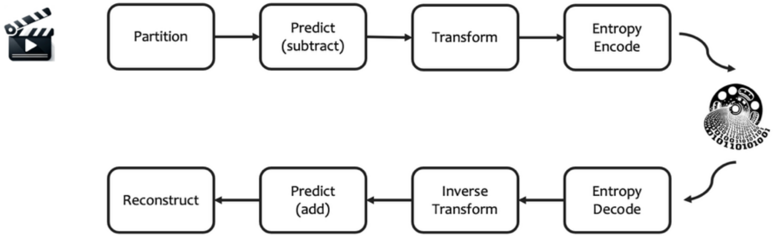


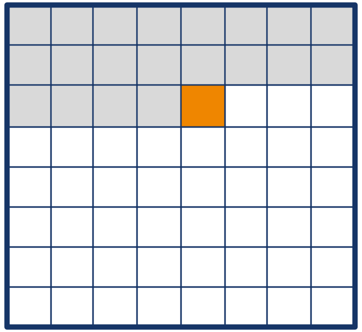
Video Coding Building Blocks





Video Coding Building Blocks





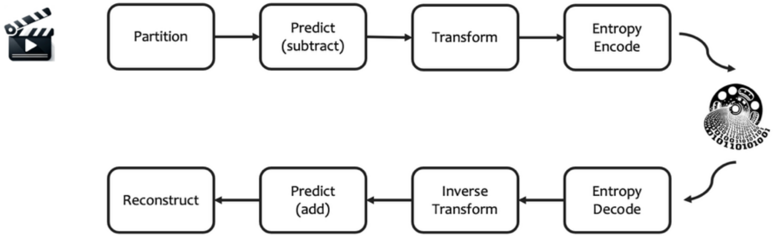
Predict

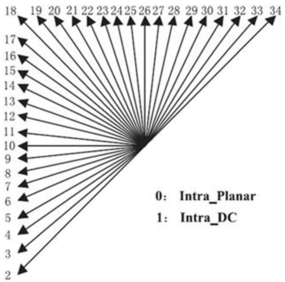
Removes redundancy by predicting the current block from

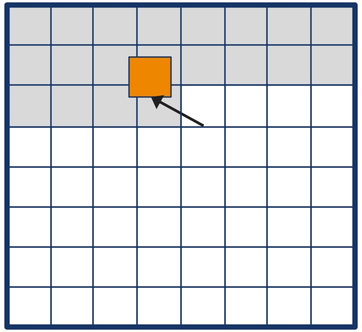
(a) *itself* (**intra**) or
(b) *previous/future* frames (**inter**)

Only the difference ("residual") is coded.

Video Coding Building Blocks







Intra coding

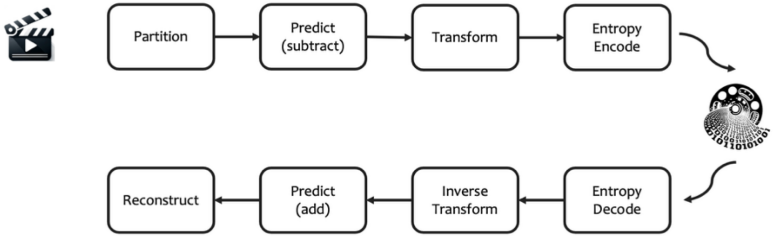
Predict

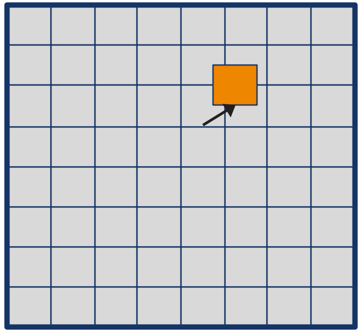
Removes redundancy by predicting the current block from

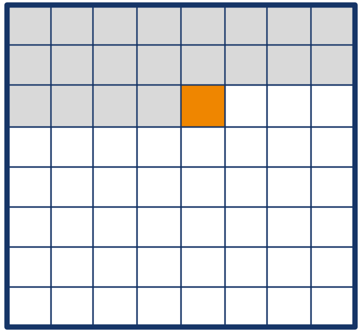
(a) *itself* (**intra**) or
(b) *previous/future* frames (**inter**)

Only the difference (“residual”) is coded.

Video Coding Building Blocks







Inter coding

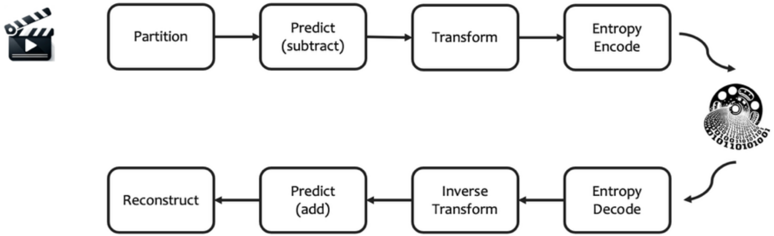
Predict

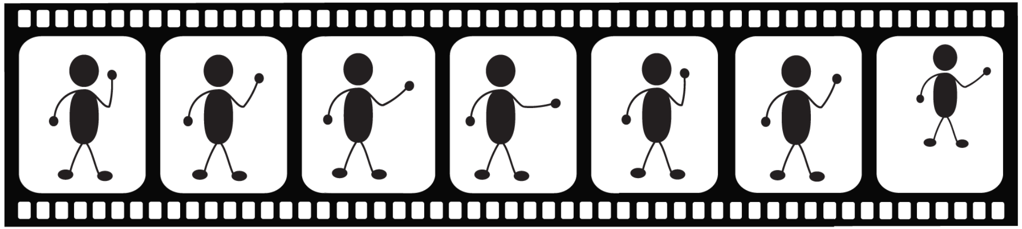
Removes redundancy by predicting the current block from

(a) *itself* (**intra**) or
(b) *previous/future* frames (**inter**)

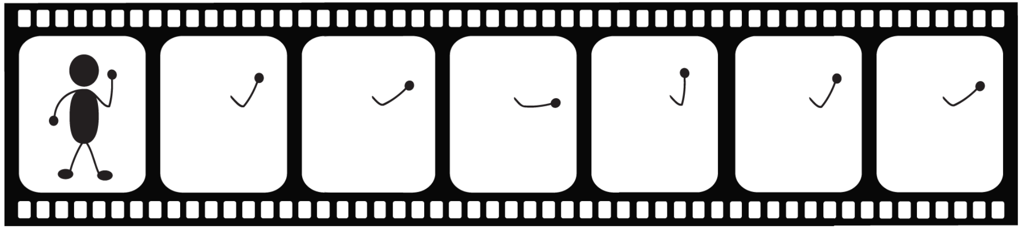
Only the difference ("residual") is coded.

Video Coding Building Blocks



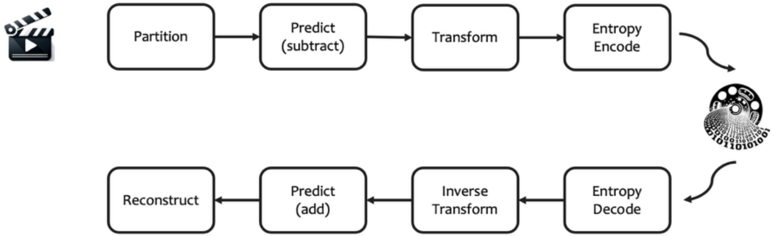


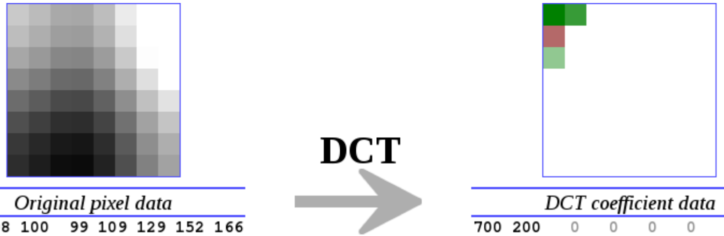
Intraframe Compression
Every frame is encoded Individually



Interframe Compression
Only the differences between frames are encoded for each group of frames

Video Coding Building Blocks





Original pixel data

114	108	100	99	109	129	152	166
109	102	95	94	104	124	146	161
99	93	85	84	94	114	137	151
86	80	72	71	82	102	124	138
73	66	58	57	68	88	110	125
60	53	46	45	55	75	97	112
50	43	36	35	45	65	88	102
45	38	31	30	40	60	82	97

DCT coefficient data

700	200	0	0	0	0	0	0
-150	0	0	0	0	0	0	0
110	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

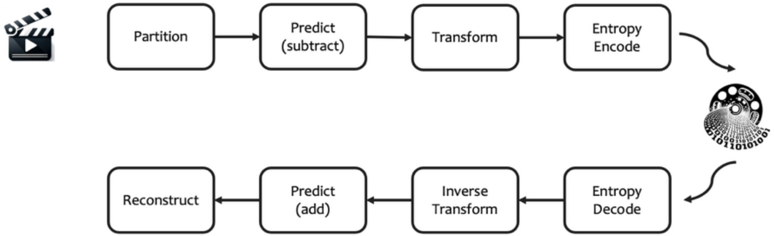
Transform

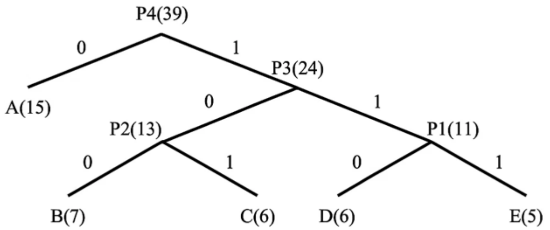
Packs residual energy into a few coefficients so most can later be quantized to zero.

UNIVERSITÄT KLAGENFURT

ATHENA

Video Coding Building Blocks





Symbol	Count	Code	Subtotal (# of bits)
A	15	0	15
B	7	100	21
C	6	101	18
D	6	110	18
E	5	111	15

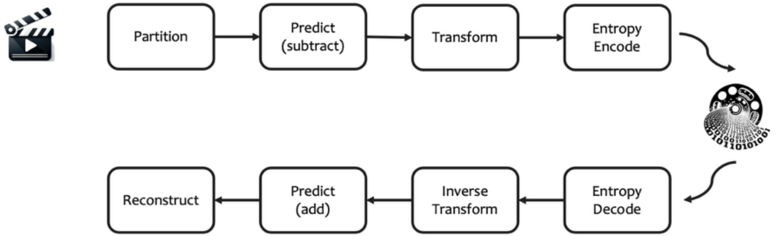
TOTAL (# of bits): 87

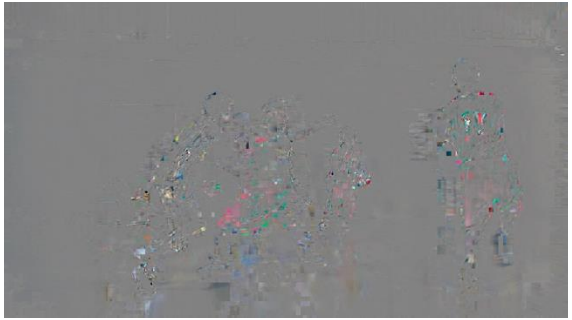
List: A(15),B(7),C(6),D(6),E(5)

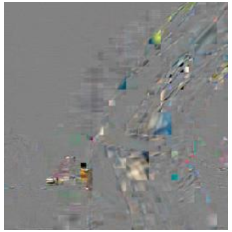
Entropy Encode

Turns runs of quantized coefficients, motion data, headers, etc. into a compact bitstream.

Video Coding Building Blocks





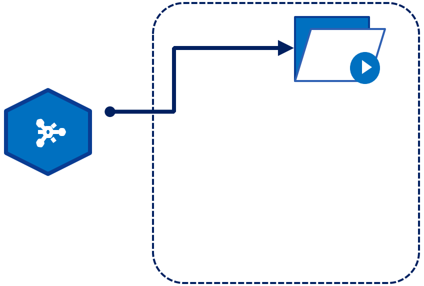
 =  + 

reconstructed | predicted | residual

UNIVERSITÄT KLAGENFURT

ATHENA
Christian Doppler (CD) Laboratory

Video Codec Timeline



AVC (2003)
The most widely adopted video codec developed by JVT: VCEG +MPEG

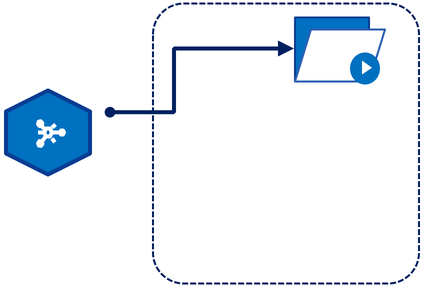
VP9 (2013)
A royalty-free codec developed by google and it is widely used in YouTube videos

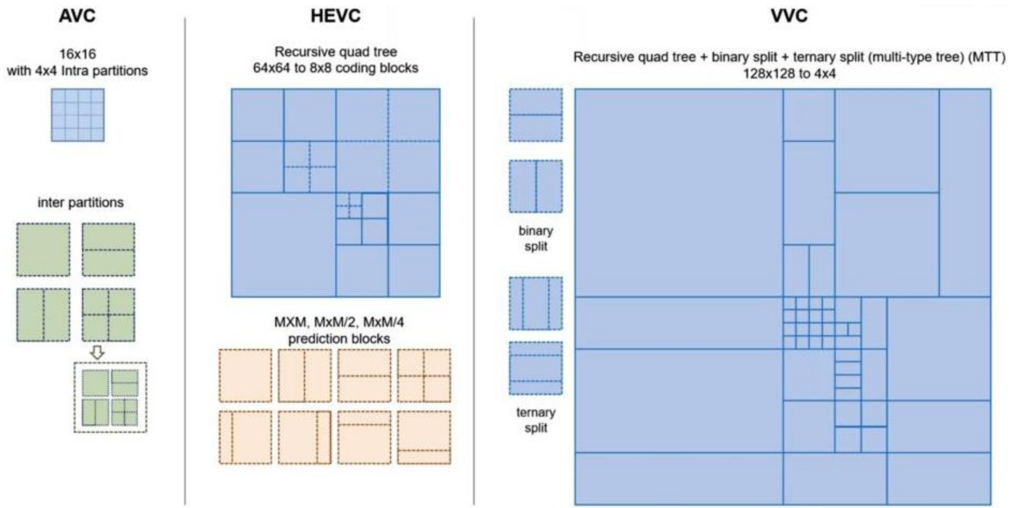
HEVC (2013)
It achieves up to 50% bitrate savings compared to AVC developed by JVT-VC: VCEG +MPEG

AV1 (2018)
A royalty-free codec developed by AOM as a successor to VP9, achieving more than 30% bitrate savings compared to VP9

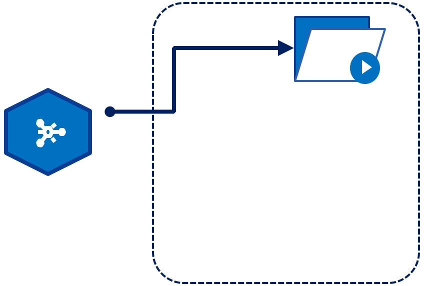
VVC (2020)
It achieve up to 50% bitrate savings compared to HEVC, known for its versatility. It was developed by JVET: VCEG+MPEG

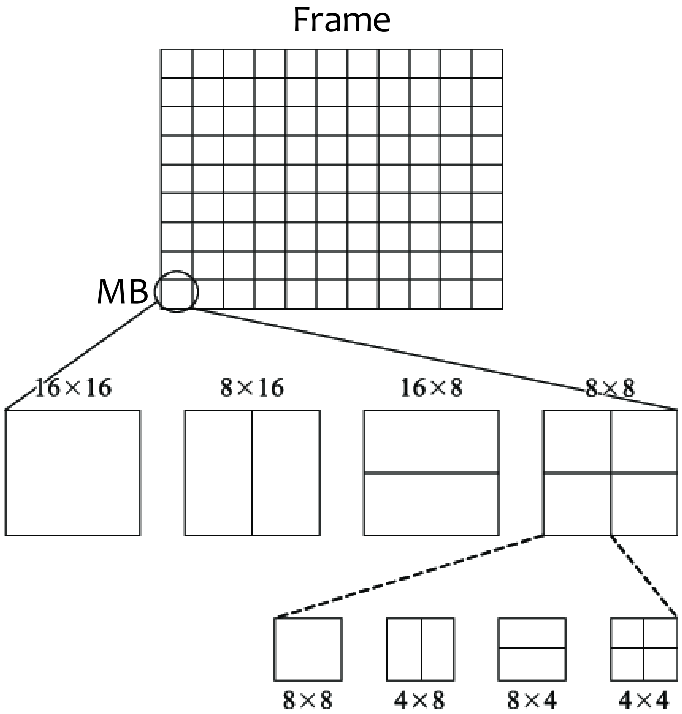
Video Codec Block Partitioning





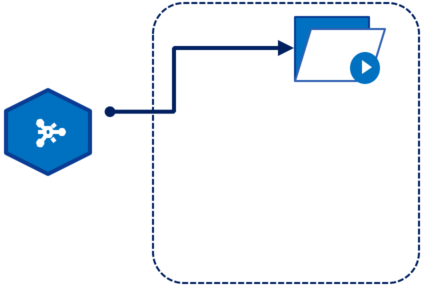
Video Codec Block Partitioning (AVC)

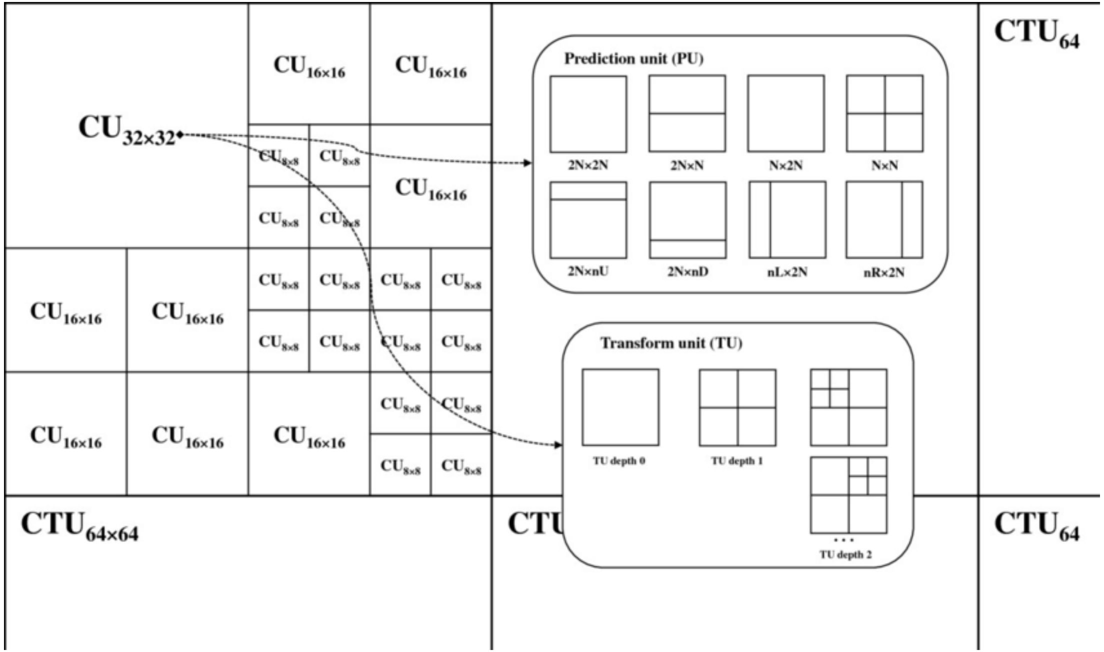




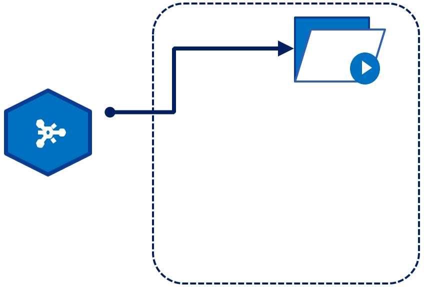
Mode	Partition Types
Inter	16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4
Intra	16×16 (Intra_16x16), 4×4 (Intra_4x4), 8×8 (Intra_8x8 in high profile)
Transform	4×4 (default), 8×8 (high profile)

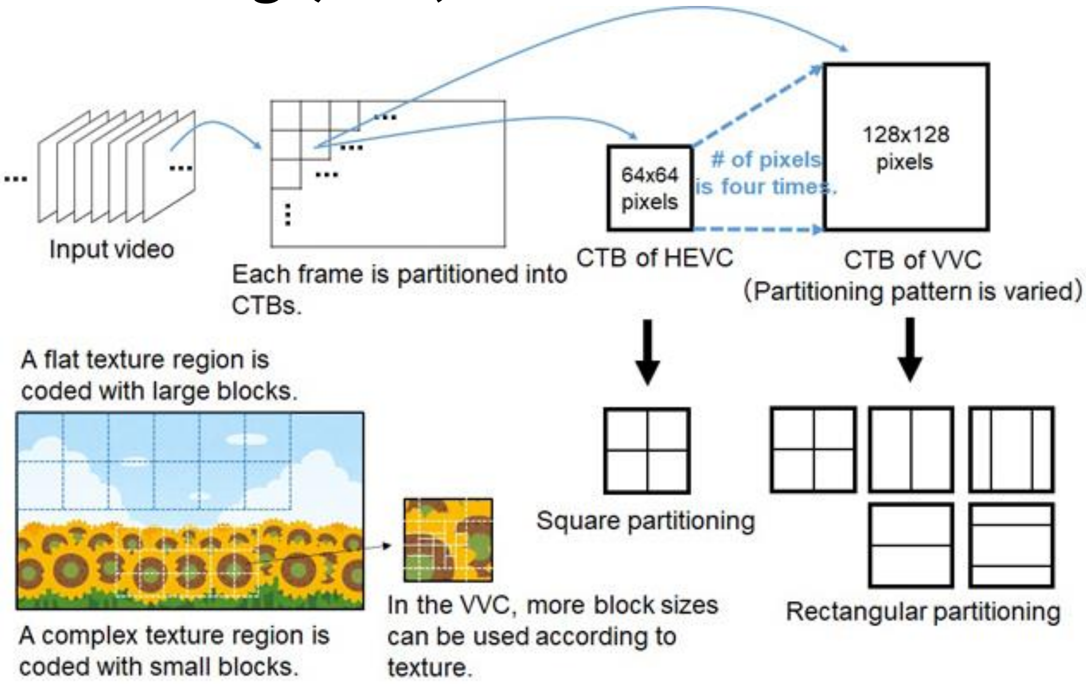
Video Codec Block Partitioning (HEVC)



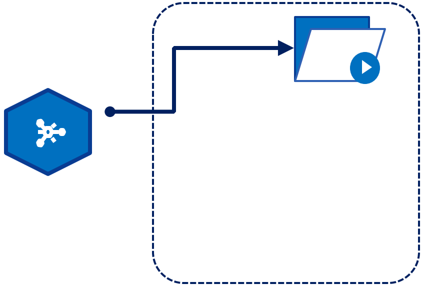


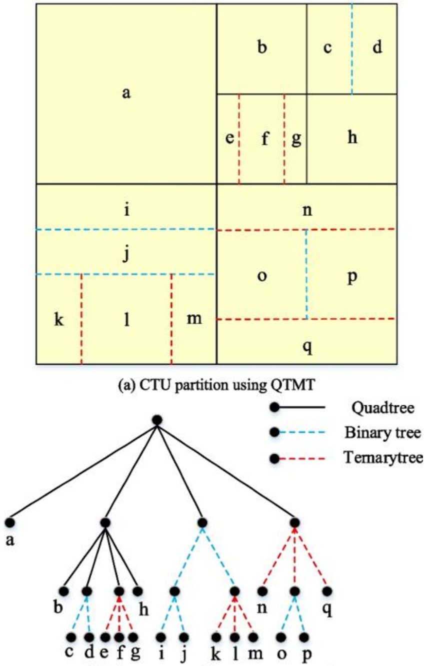
Video Codec Block Partitioning (VVC)





Video Codec Block Partitioning (VVC)

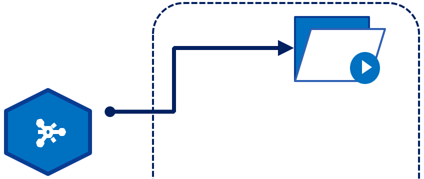


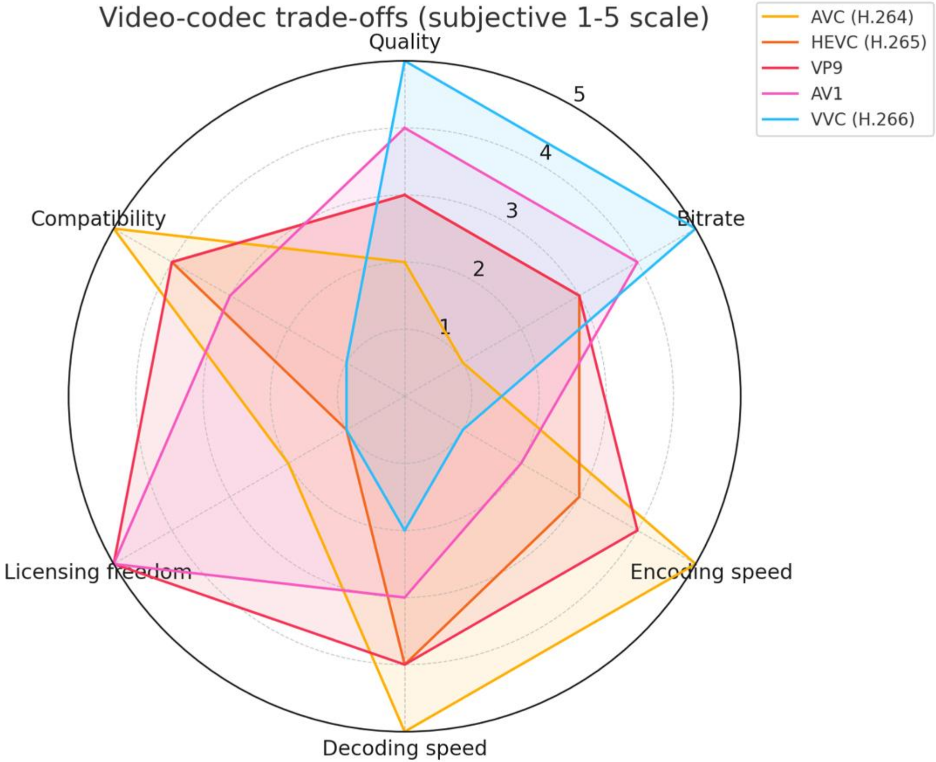


(a) CTU partition using QTMT

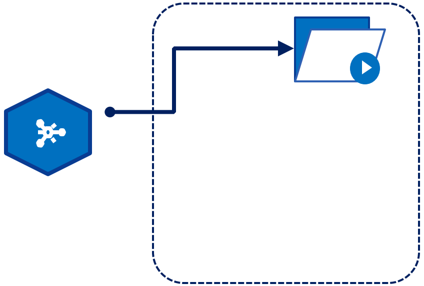
(b) The corresponding tree representation

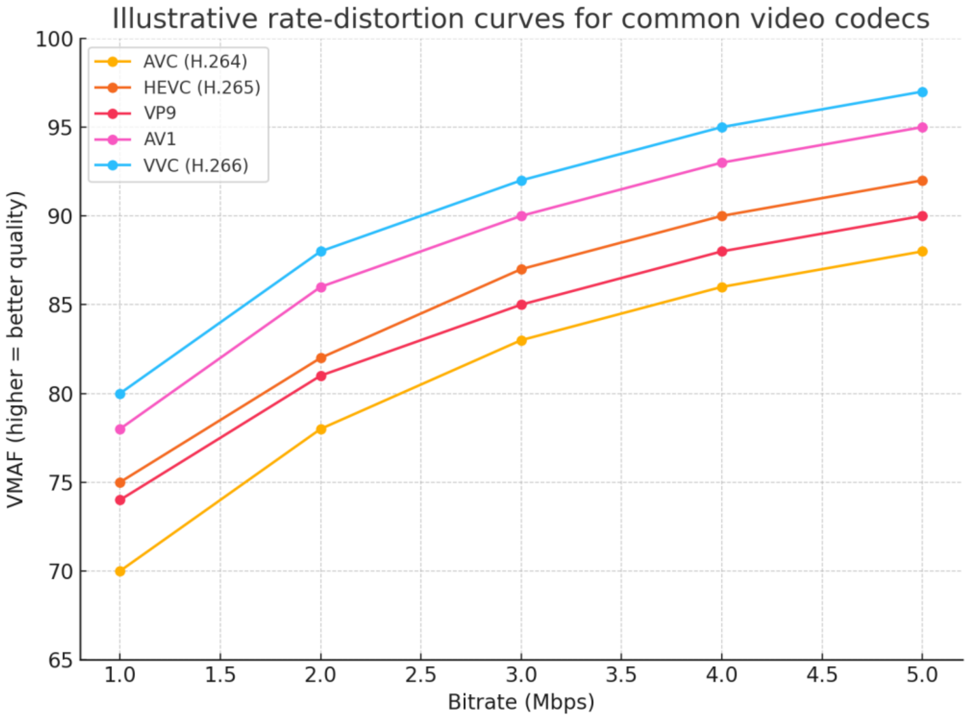
Video Codec Evaluation



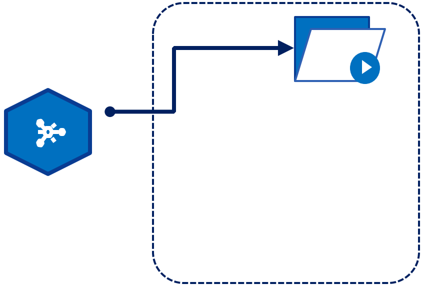


Video Codec Evaluation (compression efficiency)





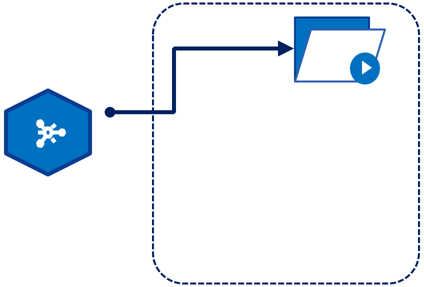
Video Codec Evaluation (Implementation)



LIST OF VIDEO CODECS FOR VARIOUS VIDEO CODING STANDARDS: "E" DENOTES ENCODERS ONLY, "D" DENOTES DECODERS ONLY, AND "ED" INDICATES THAT BOTH ENCODER AND DECODER ARE AVAILABLE.

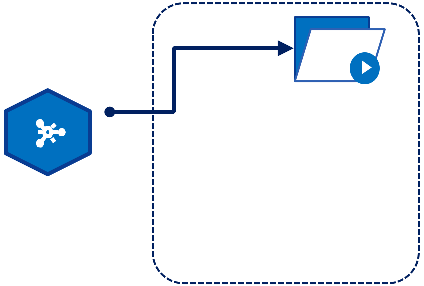
Codec	Software Implementation	Hardware Implementation
H.264	x264 [e], Nero Digital [ed] MainConcept [ed], OpenH264 [ed]	Intel Quick Sync [ed], AMD Video Coding Engine (VCE) [e] AMD Video Core Next (VCN) [ed], Nvidia NVENC [e], Nvidia NVDEC [d], Nvidia PureVideo [d]
H.265	x265 [e], OpenHEVC [ed], Turing [ed], Aurora5 [ed]	Intel Quick Sync Video [ed], ARM (Media Foundation) [ed] AMD Video Core Next (VCN) [ed], Nvidia NVENC [e], Nvidia NVDEC [d]
VP9	libvpx [ed], SVT-VP9 [e], ffvp9 [d], and Eve [e]	Intel Quick Sync Video [ed], AMD Video Coding Next (VCE) [d], Nvidia NVENC (e), Nvidia NVDEC (d)
AV1	Libaom [ed], rav1e [e], SVT-AV1 [ed] dav1d [d], Cisco AV1 [ed], libgav1 [d]	Intel Quick Sync Video AMD Video Core Next (VCN) [ed], Nvidia NVENC (e), Nvidia NVDEC (d)
VVC	VVenC [e], VVdeC [d], uvg266 [e], openVVC [d]	-

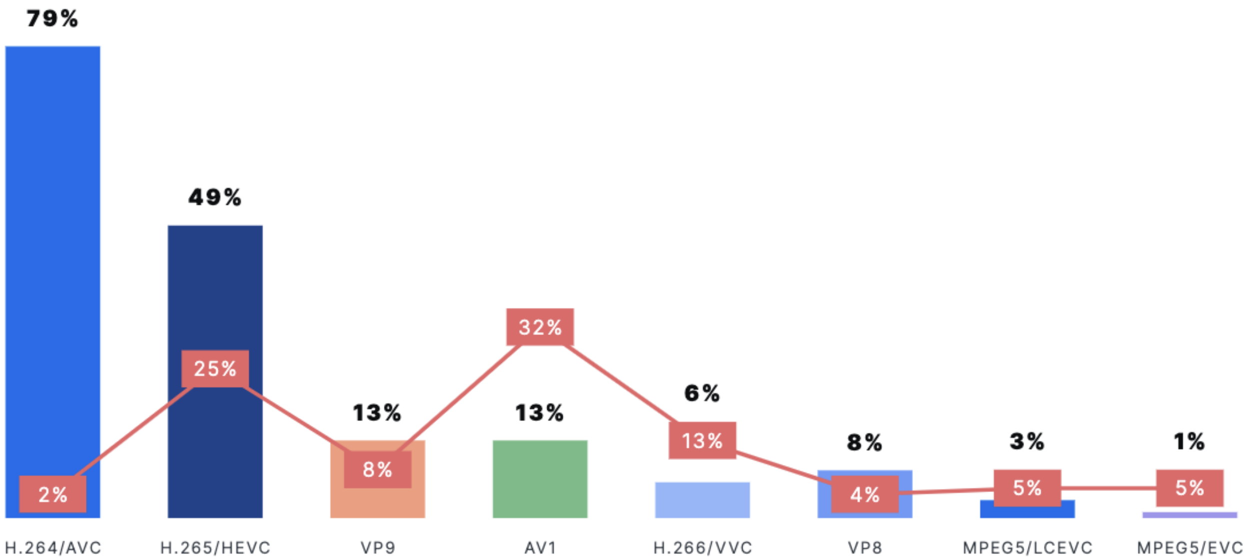
Video Codec Evaluation (Compatibility)

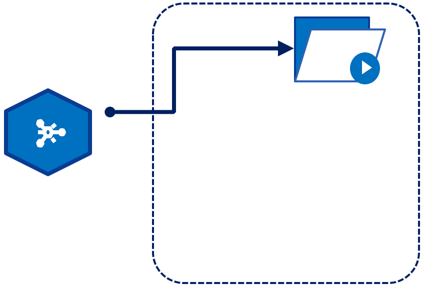


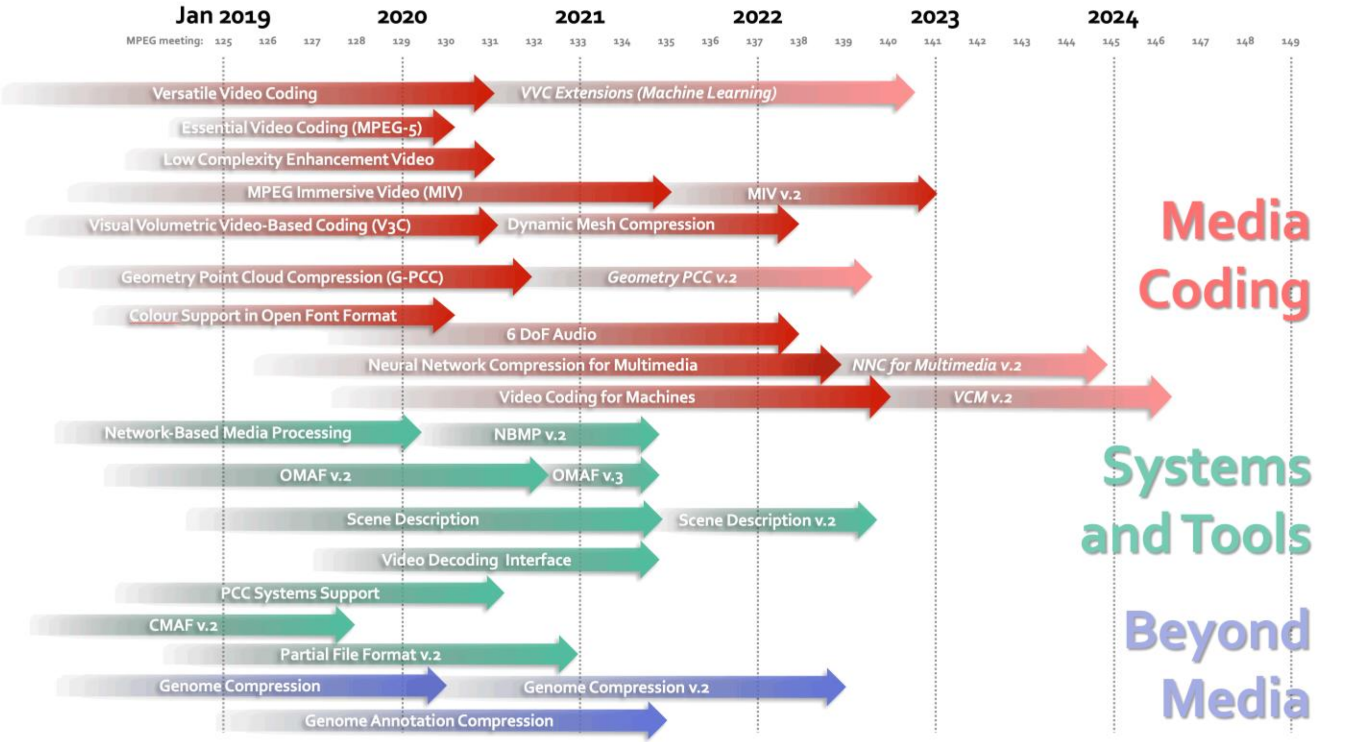
Codec \ Browser	Chrome	Safari	Firefox	Edge
H.264	Version 4 and onwards	Version 3.2 and onwards	Version 35 and onwards	Version 12 and onwards
H.265	Not Supported by default	Version 13 and onwards	Not supported	Not Supported by default
VP9	Version 70 and onwards	Not supported	Version 67 and onwards	Not supported
AV1	Version 70 and onwards	Not supported	Version 76 and onwards	Not supported

Video Codec Evaluation (Compatibility)



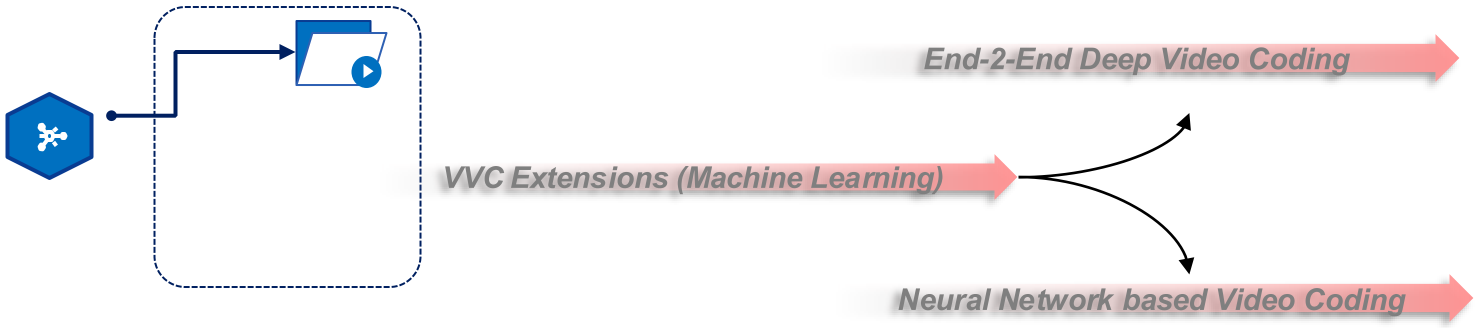




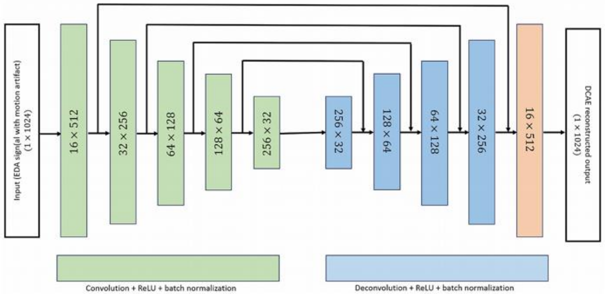


Jan 2019
2020
2021
2022
2023
2024
MPEG meeting: 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149
Versatile Video Coding
VVC Extensions (Machine Learning)
Essential Video Coding (MPEG-5)
Low Complexity Enhancement Video
MPEG Immersive Video (MIV)
MIV v.2
Visual Volumetric Video-Based Coding (V3C)
Dynamic Mesh Compression
Geometry Point Cloud Compression (G-PCC)
Geometry PCC v.2
Colour Support in Open Font Format
6 DoF Audio
Neural Network Compression for Multimedia
NNC for Multimedia v.2
Video Coding for Machines
VCM v.2
Network-Based Media Processing
NBMP v.2
OMAF v.2
OMAF v.3
Scene Description
Scene Description v.2
Video Decoding Interface
PCC Systems Support
CMAF v.2
Partial File Format v.2
Genome Compression
Genome Compression v.2
Genome Annotation Compression
Media Coding
Systems and Tools
Beyond Media

Next-generation Video Coding Standard

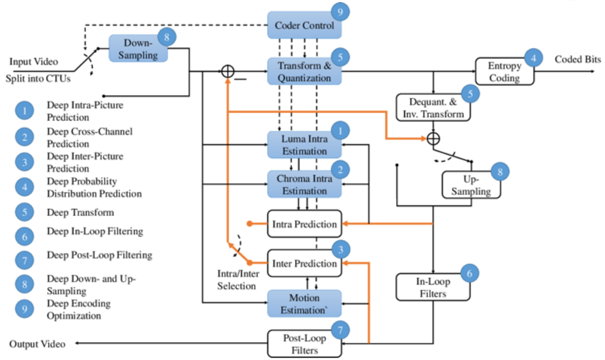


VVC Extensions (Machine Learning)



End-2-End Deep Video Coding

Neural Network based Video Coding



Next-generation Video Coding Standard

- 🧩 **Keeps VVC syntax, swaps tools** – small CNNs for in-loop filtering, intra prediction, motion refinement, RDO decisions
- 📉 **-12–25 % BD-rate over VVC** with ECM-14 configuration; gains vary by content and how many NN tools are enabled
- ⚙️ **Complexity bump, but ASIC-friendly** – ≤ 64 kB weights per model; extra NN accelerator block plus existing video core
- ↔ **Backward path** – bit-stream largely VVC-compatible; legacy decoders can ignore the extra NAL/SEI units

Next-generation Video Coding Standard

In-Loop Filter

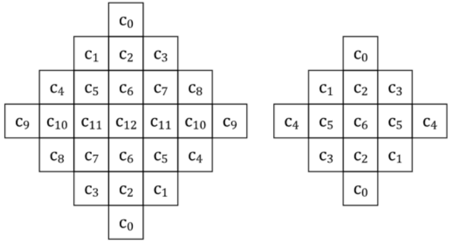
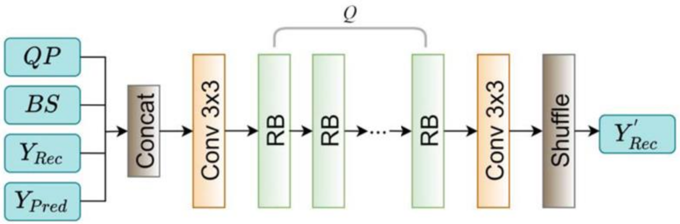
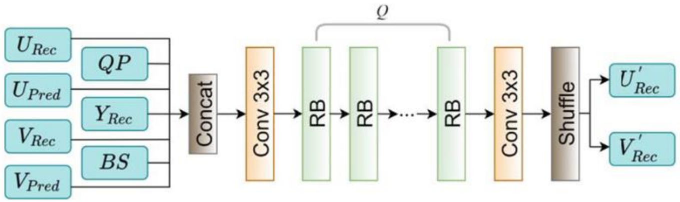


Fig. 2. ALF filter shapes.



(a) Network structure for luma component.



(b) Network structure for chroma components.

Next-generation Video Coding Standard

- **Pure learned codec** – transformer/CNN auto-encoder replaces the entire hybrid pipeline; only a thin bit-stream wrapper remains
- 📉 **-40–60 % BD-rate vs. VVC** on HD/4K test sets (best lab numbers, 2025)
- 🚀 **Needs heavy AI silicon** – real-time today only on datacentre GPUs or flagship NPUs; both encoder *and* decoder are GPU-class workloads
- 🔄 **Training-data now part of conformance** – models, checkpoints and dataset IDs must be signalled in the spec
- 📅 **Standardisation** – JVET "EE-1" test model v10; first full neural CfE expected late-2025; commercial deployment realistically ≥ 2028

Next-generation Video Coding Standard

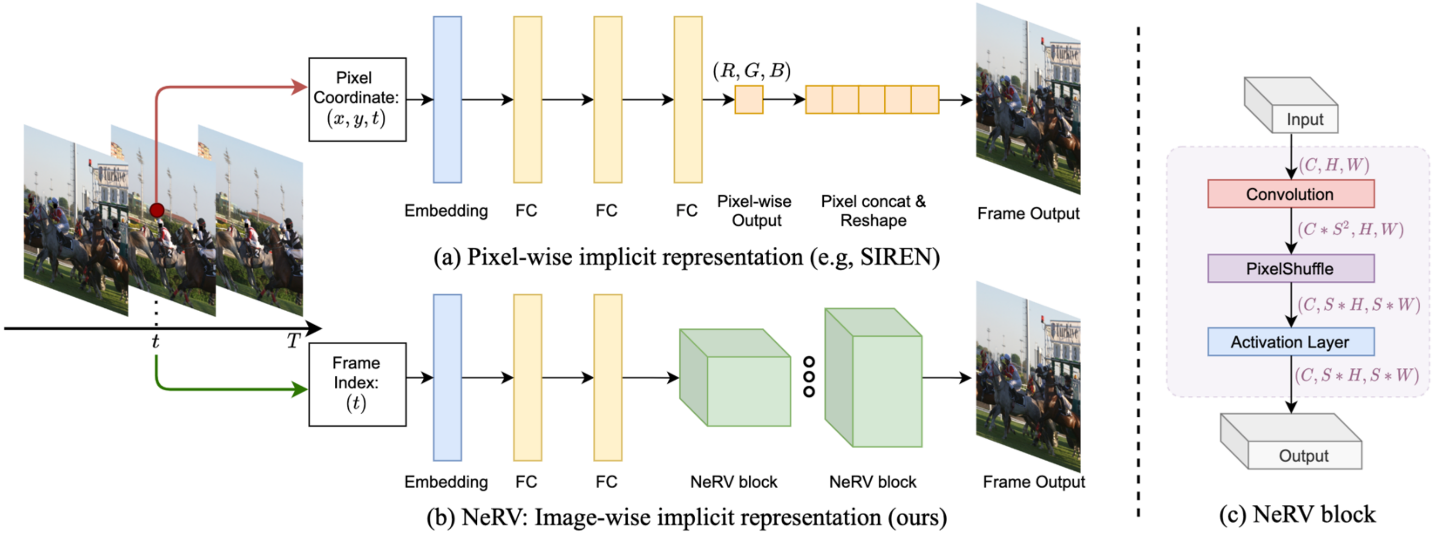


Figure 2: **(a) Pixel-wise** implicit representation taking pixel coordinates as input and use a simple MLP to output pixel RGB value **(b) NeRV: Image-wise** implicit representation taking frame index as input and use a MLP + ConvNets to output the whole image. **(c) NeRV block** architecture, upscale the feature map by S here.

Next-generation Video Coding Standard

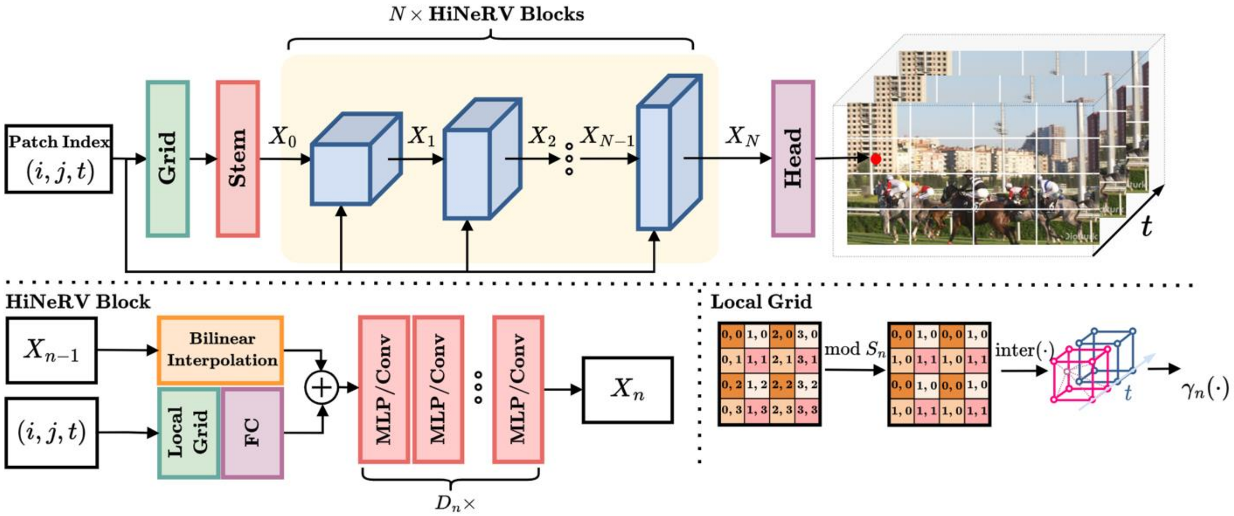


Figure 2: Top: The HiNeRV architecture. Bottom left: The HiNeRV block. HiNeRV block take feature maps X_{n-1} and patch index (i, j, t) as input, upsample the feature maps, enhances it with the hierarchical encoding, then computes the transformed maps X_n. Bottom right: The local grid. In HiNeRV, the hierarchical encoding is computed by performing interpolation from the local grid, where the modulo of the coordinates is being used.

HiNeRV: Video Compression with Hierarchical Encoding-based Neural Representation, NIPS 2023

Next-generation Video Coding Standard

Table 2: Video representation results with the UVG dataset [38] (for S/M/L scales). Results are in PSNR. FPS is the encoding/decoding rate.

Model	Size	MACs	FPS	Beauty	Bosph.	Honey.	Jockey	Ready.	Shake.	Yacht.	Avg.
NeRV	3.31M	227G	**32.4**/90.0	32.83	32.20	38.15	30.30	23.62	33.24	26.43	30.97
E-NeRV	3.29M	230G	20.7/75.9	33.13	33.38	38.87	30.61	24.53	34.26	26.87	31.75
PS-NeRV	3.24M	538G	14.7/42.6	32.94	32.32	38.39	30.38	23.61	33.26	26.33	31.13
HNeRV	3.26M	175G	24.6/**93.4**	33.56	35.03	39.28	31.58	25.45	34.89	28.98	32.68
FFNeRV	3.40M	228G	19.0/49.3	33.57	35.03	38.95	31.57	25.92	34.41	28.99	32.63
HiNeRV	3.19M	181G	10.1/35.5	**34.08**	**38.68**	**39.71**	**36.10**	**31.53**	**35.85**	**30.95**	**35.27**
NeRV	6.53M	228G	**32.0/90.1**	33.67	34.83	39.00	33.34	26.03	34.39	28.23	32.78
E-NeRV	6.54M	245G	20.5/74.6	33.97	35.83	39.75	33.56	26.94	35.57	28.79	33.49
PS-NeRV	6.57M	564G	14.6/42.0	33.77	34.84	39.02	33.34	26.09	35.01	28.43	32.93
HNeRV	6.40M	349G	20.1/68.5	33.99	36.45	39.56	33.56	27.38	35.93	30.48	33.91
FFNeRV	6.44M	229G	18.9/49.3	33.98	36.63	39.58	33.58	27.39	35.91	30.51	33.94
HiNeRV	6.49M	368G	8.4/29.1	**34.33**	**40.37**	**39.81**	**37.93**	**34.54**	**37.04**	**32.94**	**36.71**
NeRV	13.01M	230G	**31.7/89.8**	34.15	36.96	39.55	35.80	28.68	35.90	30.39	34.49
E-NeRV	13.02M	285G	21.0/68.1	34.25	37.61	39.74	35.45	29.17	36.97	30.76	34.85
PS-NeRV	13.07M	608G	14.1/41.4	34.50	37.28	39.58	35.34	28.56	36.51	30.28	34.61
HNeRV	12.87M	701G	15.6/52.7	34.30	37.96	39.73	35.47	29.67	37.16	32.31	35.23
FFNeRV	12.66M	232G	18.4/49.3	34.28	38.48	39.74	36.72	30.75	37.08	32.36	35.63
HiNeRV	12.82M	718G	5.5/19.9	**34.66**	**41.83**	**39.95**	**39.01**	**37.32**	**38.19**	**35.20**	**38.02**

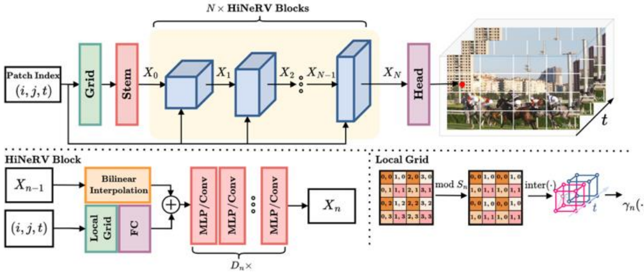
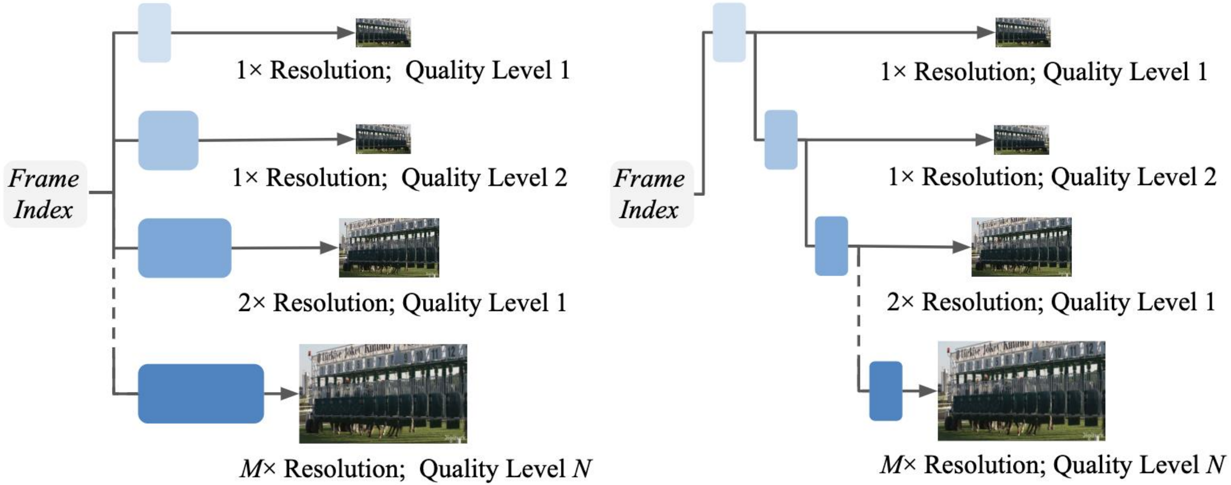


Figure 2: Top: The HiNeRV architecture. Bottom left: The HiNeRV block. HiNeRV block take feature maps X_{n-1} and patch index (i, j, t) as input, upsample the feature maps, enhances it with the hierarchical encoding, then computes the transformed maps X_n. Bottom right: The local grid. In HiNeRV, the hierarchical encoding is computed by performing interpolation from the local grid, where the modulo of the coordinates is being used.

Next-generation Video Coding Standard



(a) Single-stream NeRV in other methods. (b) Multi-stream NeRV in our proposed NSVC.

Next-generation Video Coding Standard

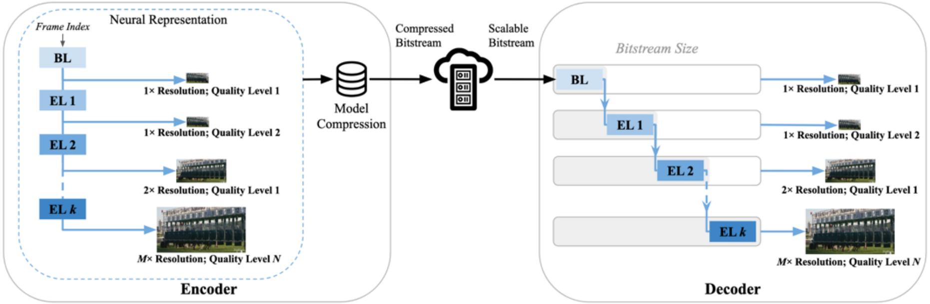
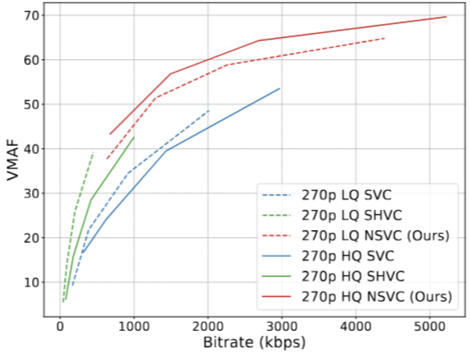


Fig. 2: The proposed Neural representations for Scalable Video Coding (NSVC) framework.



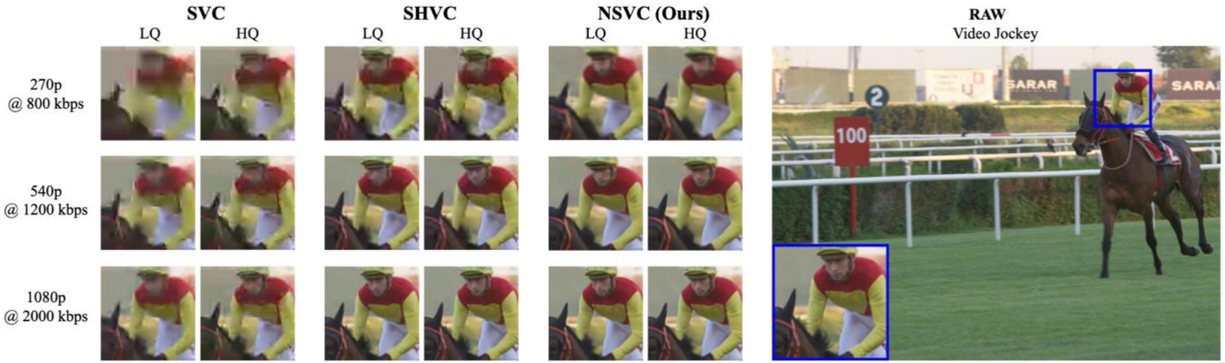
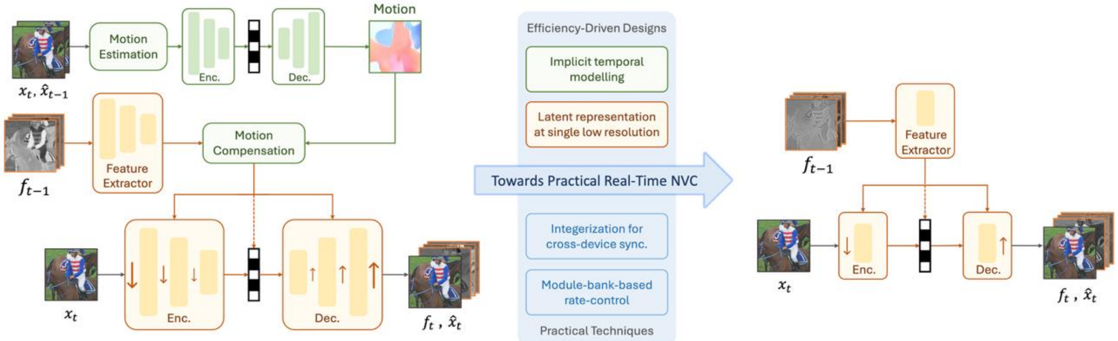


Fig. 5: Visual comparison of among cropped reconstructed frames from SVC, SHVC, and NSVC (Ours).

Next-generation Video Coding Standard



(a) Paradigm of motion-based NVCs

(b) Proposed practical real-time NVC

Figure 2. Paradigm shift. To enhance efficiency, we eliminate explicit motion-related modules and adopt implicit temporal modeling. We also propose learning latent representations at a single low resolution, replacing the traditional progressive downsampling approach. Additionally, DCVC-RT supports integerization for cross-device consistency and incorporates a module-bank-based rate-control mechanism.

Video Codecs	BD-Rate against x265 on UVG ↓	1080p Coding Speed / FPS ↑			Cross Device	Rate Control
		Device	Enc.	Dec.		
Traditional Codecs						
NVEnc-HEVC	-7.9%	RTX 4090	418	502	✓	✓
VTM-17.0	-62.0%	AMD EPYC 7V13 Processor	0.01	23.6	✓	✓
ECM-11.0	-69.4%	AMD EPYC 7V13 Processor	0.002	3.4	✓	✓
Neural Codecs						
MobileNVC	50.8%	Snapdragon 8 Gen 2	3	38.9	✓	
ELF-VC	-26.1%	Titan V	10	18		✓
C3	-29.0%	A100	0.0004	15.6		
DHVC-2.0	-56.9%	RTX 3090	4.3	7.1		✓
DCVC-FM	-69.8%	A100	5.0	5.9		✓
DCVC-RT	**-70.9%**	RTX 2080Ti	**40**	**34**	✓	✓
		RTX 4090	**119**	**105**		
		A100	**125**	**113**		

Next-generation Video Coding Standard

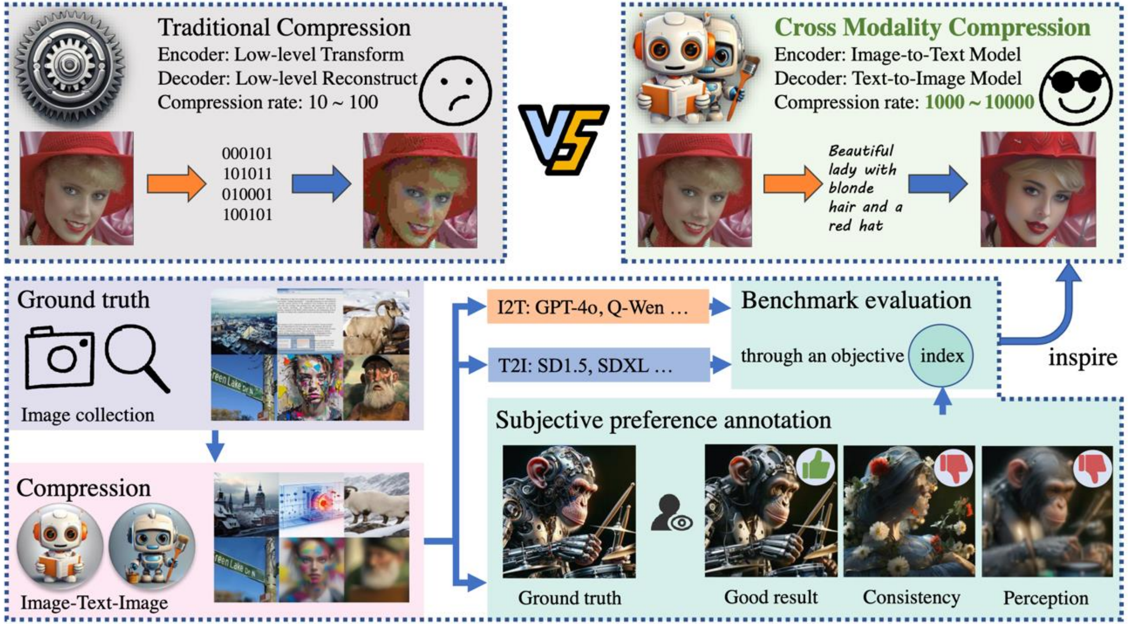


Figure 1: Overview of CMC-Bench. We demonstrate the superiority of Cross Modality Compression over traditional codecs, and subjective and objective evaluations of compression results on Consistency and Perception. This benchmark can motivate it to become the future codec paradigm.

Next-generation Video Coding Standard

Text: The I2T model converts images to text and is directly restored by the T2I model. Due to its reliance on the text modality only, this approach achieves a CR of 10,000, ideal for ELB situations.

Pixel: Each 64 × 64 blocks from ground truth are merged and quantized into one pixel. Beyond the *Text* mode, these pixels initialize the T2I process. The pixel representation is relatively compact, offering a CR of around 5,000, suitable for less rigorous ELB but higher demands on consistency.

Image: Traditional codecs are employed to compress the image, which serves as input for the T2I model for enhancement. Unlike the previous two, it omits the time-consuming I2T process by leaving the text input of the T2I model empty. This approach can achieve a CR of 1,000, suitable for ULB bandwidth but with high real-time requirements.

Full: Extending the *Image* mode, the T2I is guided by text content, encompassing the full pipeline of I2T, traditional codec, and T2I. It also has a CR of approximately 1,000, suitable for the most demanding performance scenarios.

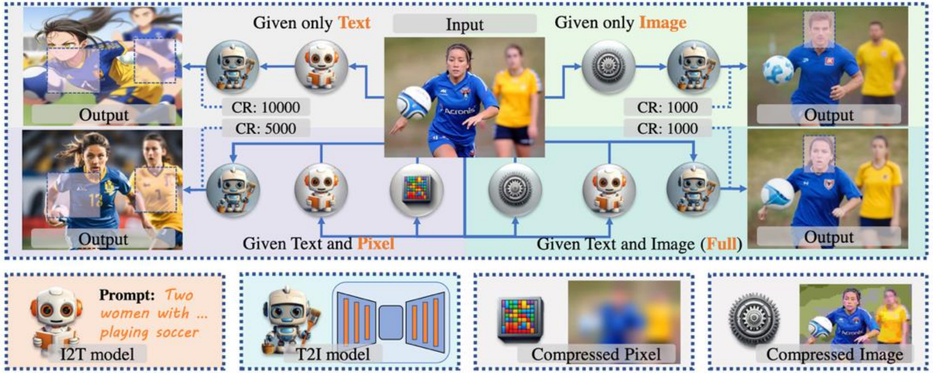
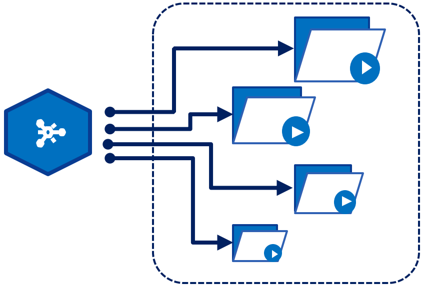


Figure 3: Illustration of 4 working modes of CMC. *Text* mode roughly reconstructs the semantic information, *Pixel* mode slightly improves low-level consistency, *Image* mode provides a similar structure towards ground truth but a different character, and *Full* mode has the best performance.

Video Coding for HAS

What is a bitrate ladder?



Vancouver 2010

	Encoding Bitrate	Resolution
Rep. #1	3.45 Mbps	1280 x 720
Rep. #2	1.95 Mbps	848 x 480
Rep. #3	1.25 Mbps	640 x 360
Rep. #4	900 Kbps	512 x 288
Rep. #5	600 Kbps	400 x 224
Rep. #6	400 Kbps	312 x 176

Sochi 2014

	Encoding Bitrate	Resolution
Rep. #1	3.45 Mbps	1280 x 720
Rep. #2	2.2 Mbps	960 x 540
Rep. #3	1.4 Mbps	960 x 540
Rep. #4	900 Kbps	512 x 288
Rep. #5	600 Kbps	512 x 288
Rep. #6	400 Kbps	340 x 192
Rep. #7	200 Kbps	340 x 192

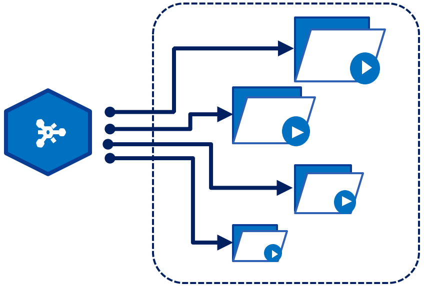
PyeongChang 2018

	Encoding Bitrate	Resolution
Rep. #1	18 Mbps	4K (60p)
Rep. #2	12.2 Mbps	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	2 Mbps	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	300 Kbps	320 x 180
Rep. #10	200 Kbps	320 x 180

Source: Vertigo MIX10, Alex Zambelli's Streaming Media Blog, Akamai, Comcast

Bitrate Ladder Construction

Fixed set of bitrates



Simple

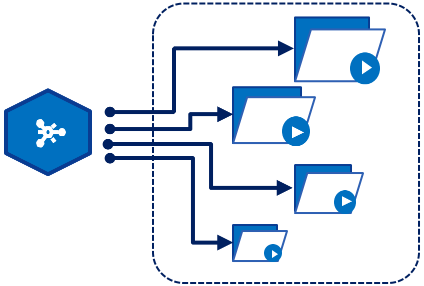
Sub-optimal

PyeongChang 2018

	Encoding Bitrate	Resolution
Rep. #1	18 Mbps	4K (60p)
Rep. #2	12.2 Mbps	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	2 Mbps	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	300 Kbps	320 x 180
Rep. #10	200 Kbps	320 x 180

Bitrate Ladder Construction

Fixed set of bitrates



Simple

Content-dependency

Network-dependency

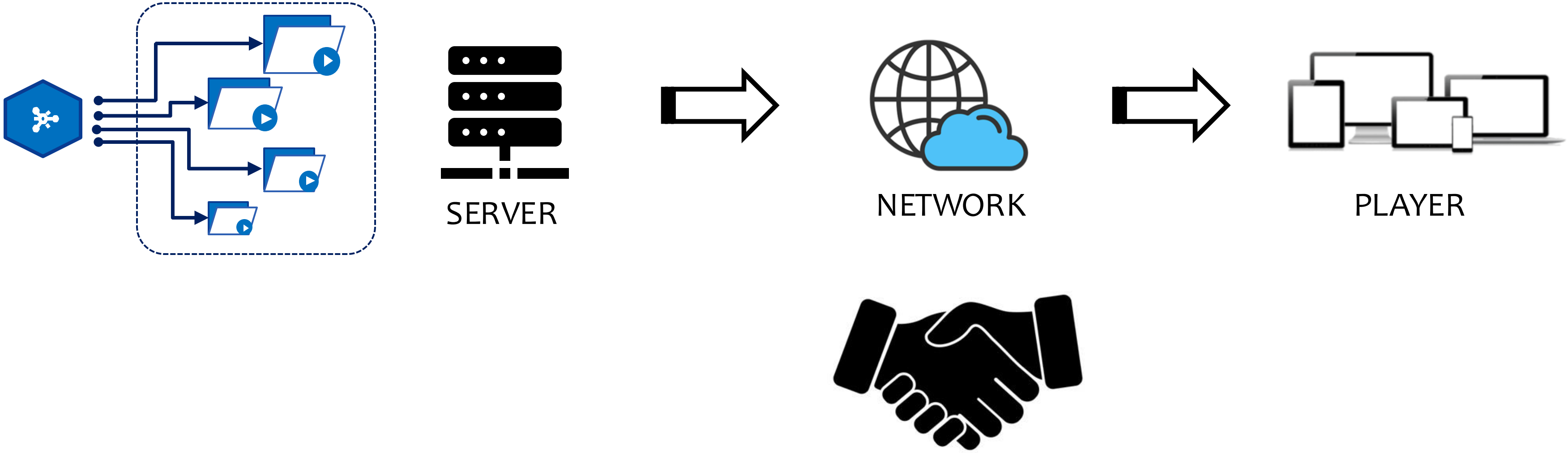
Sub-optimal

PyeongChang 2018

	Encoding Bitrate	Resolution
Rep. #1	18 Mbps	4K (60p)
Rep. #2	12.2 Mbps	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	2 Mbps	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	300 Kbps	320 x 180
Rep. #10	200 Kbps	320 x 180

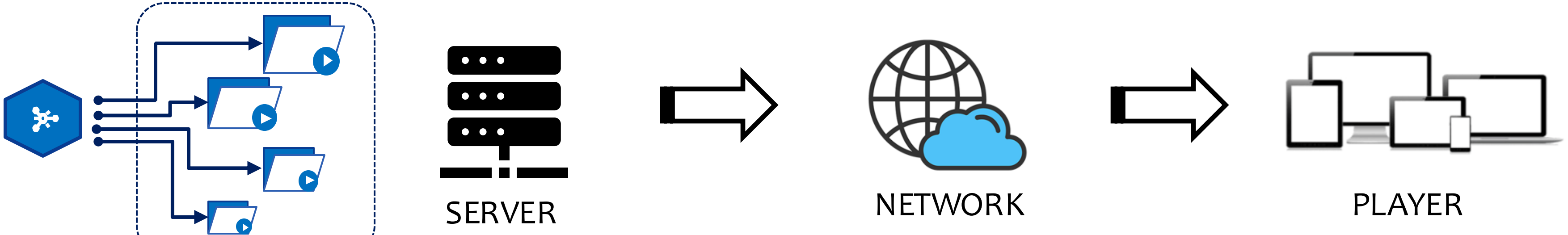
Bitrate Ladder Construction

Network-aware bitrate ladder construction

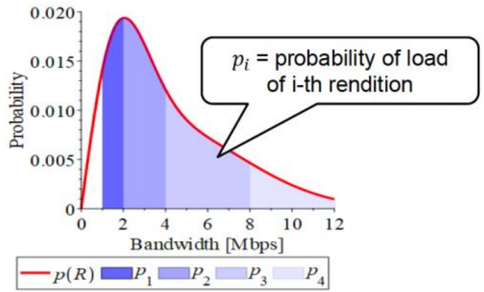


Bitrate Ladder Construction

Network-aware bitrate ladder construction



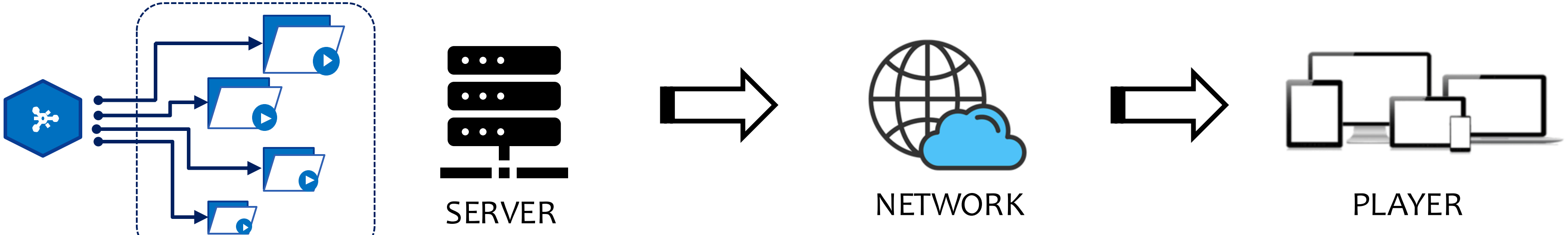
Network model $p(R)$



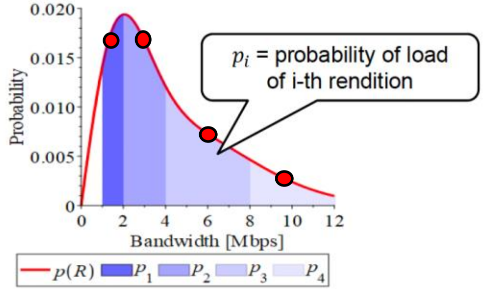
[Reznik, et al 2019, PV]

Bitrate Ladder Construction

Network-aware bitrate ladder construction



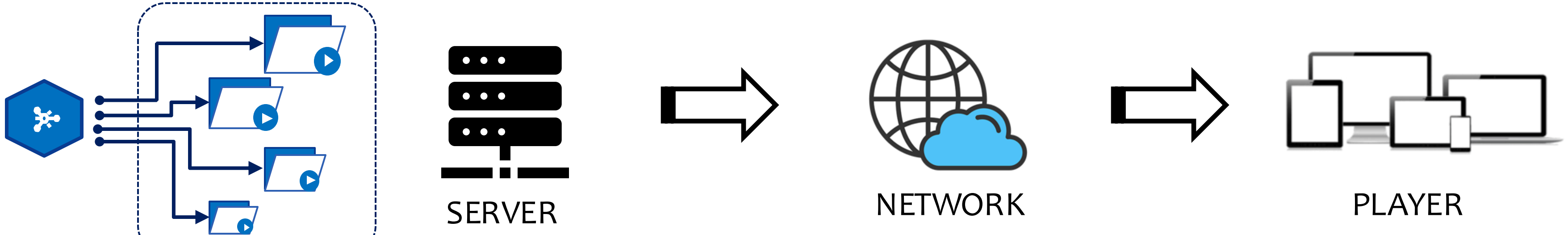
Network model $p(R)$



[Reznik, et al 2019, PV]

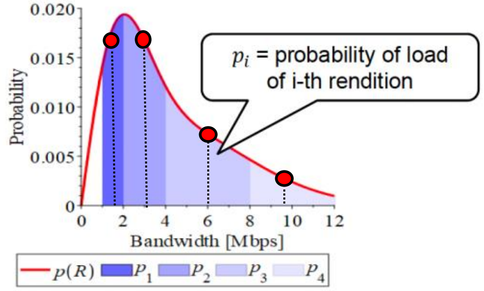
Bitrate Ladder Construction

Network-aware bitrate ladder construction



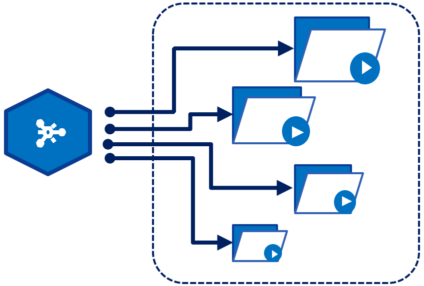
	Encoding Bitrate
Rep. #01	1 Mbps
Rep. #02	300 kbps
Rep. #03	600 kbps
Rep. #04	900 kbps

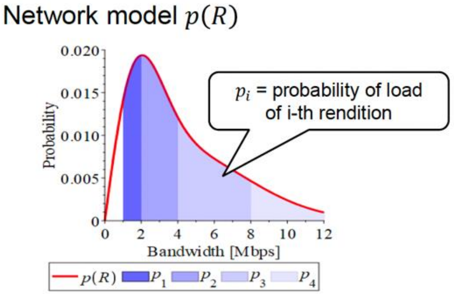
Network model $p(R)$

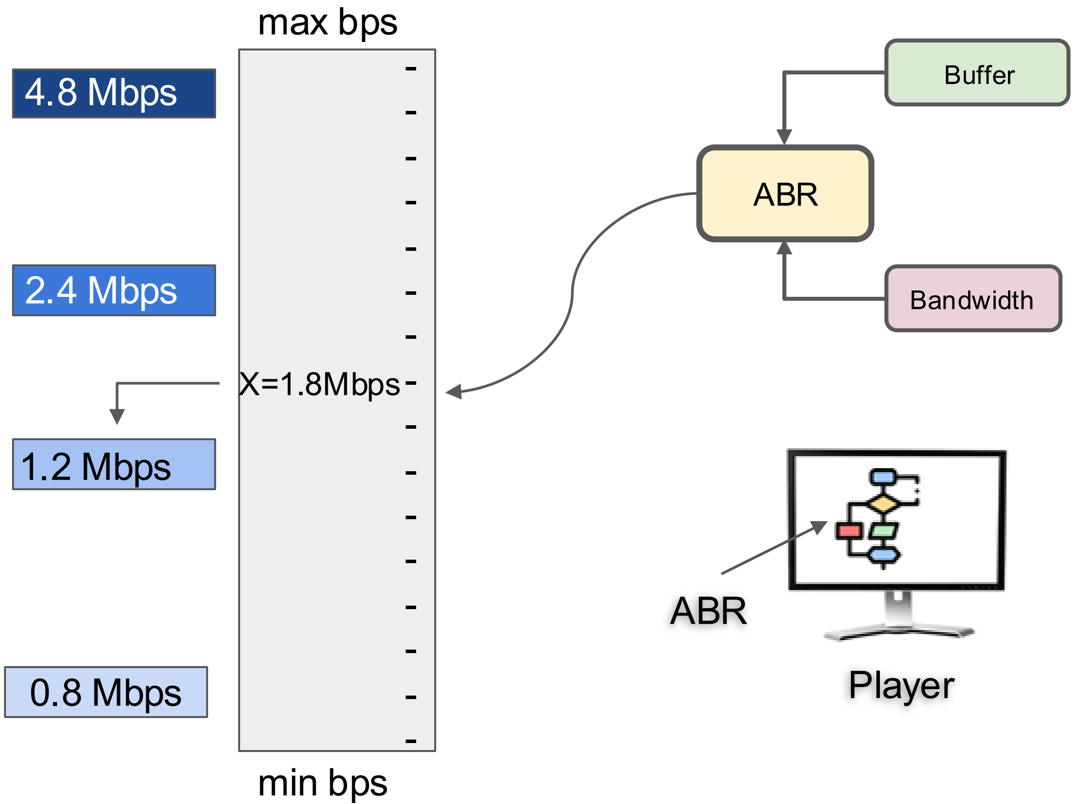


Bitrate Ladder Construction

Network-aware bitrate ladder construction (LALISA)



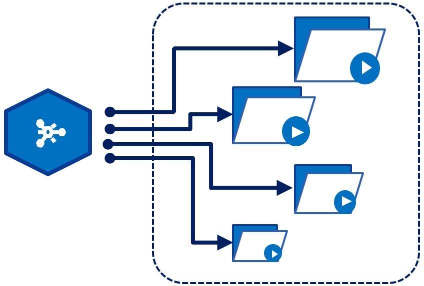




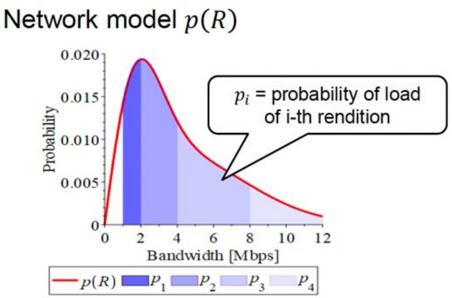
LALISA, NOMS 2023

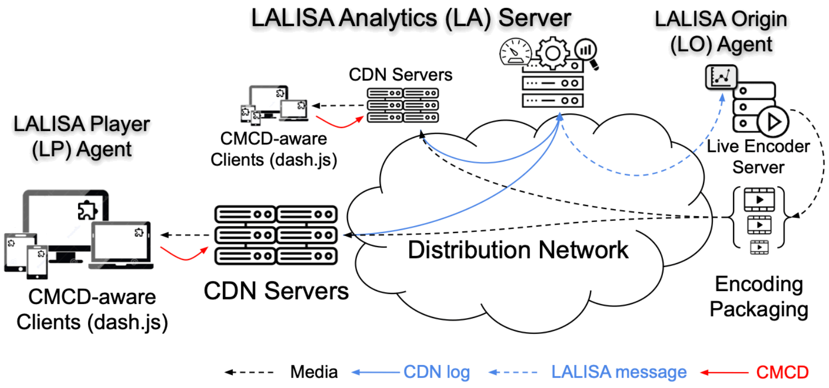
Bitrate Ladder Construction

Network-aware bitrate ladder construction (LALISA)



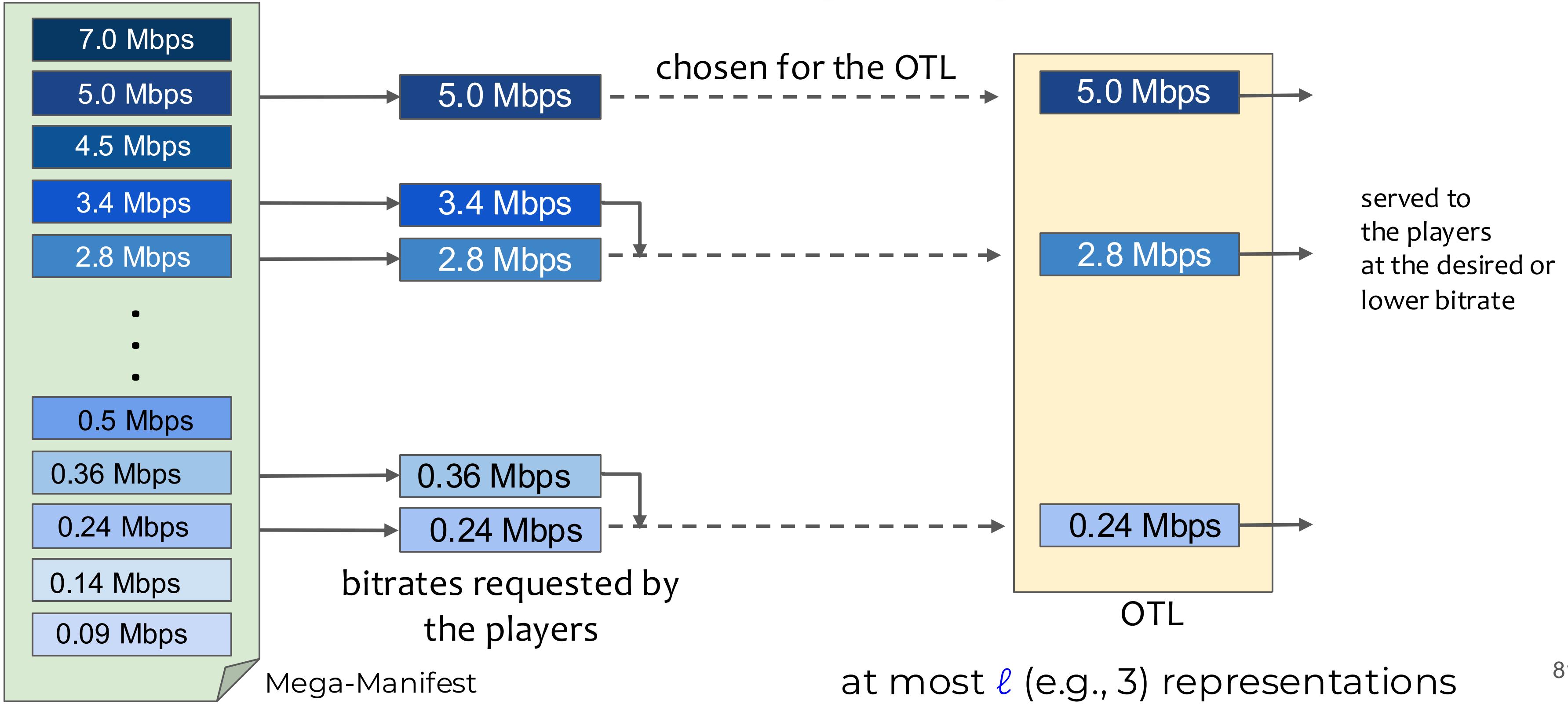
- Clients send their **desired bitrate** at each instance using **CMCD**
- The server selects bitrates to construct the bitrate ladder based on the probability of desired bitrates
- This approach requires modification in the player





Bitrate Ladder Construction

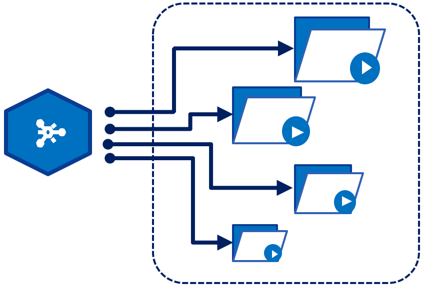
Network-aware bitrate ladder construction (ARTEMIS)



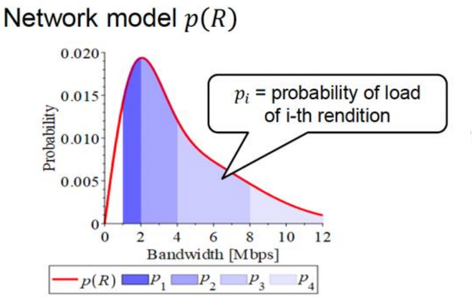
at most ℓ (e.g., 3) representations

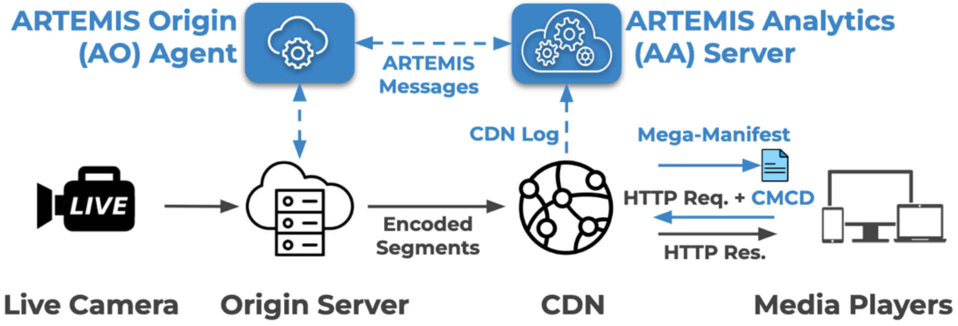
Bitrate Ladder Construction

Network-aware bitrate ladder construction (ARTEMIS)



- The concept of Mega Manifest is introduced
- The clients select their desired bitrate from the Mega Manifest
- Based on probability of the requested bitrate, the optimized ladder is constructed

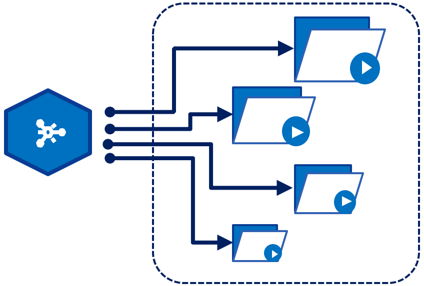




ARTEMIS, NSDI 2024

Bitrate Ladder Construction

Quality-aware bitrate ladder construction

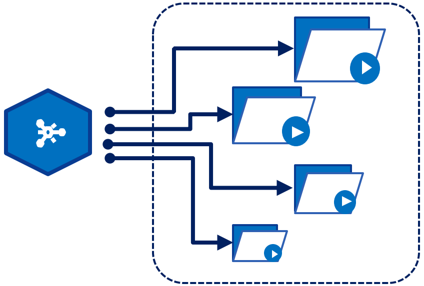


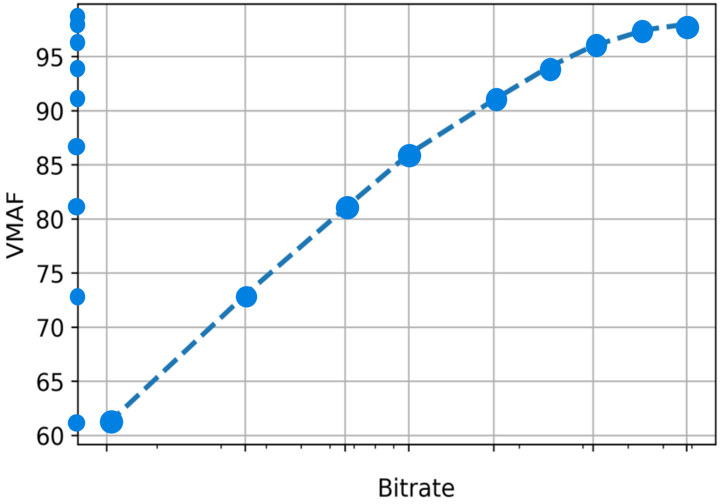
PyeongChang 2018

	Encoding Bitrate	Resolution
Rep. #1	18 Mbps	4K (60p)
Rep. #2	12.2 Mbps	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	2 Mbps	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	300 Kbps	320 x 180
Rep. #10	200 Kbps	320 x 180

Bitrate Ladder Construction

Quality-aware bitrate ladder construction



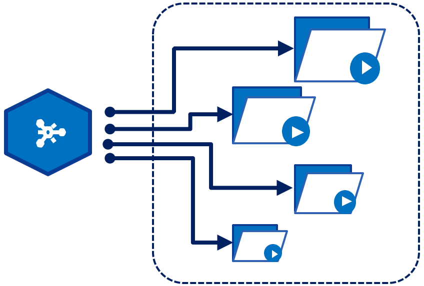


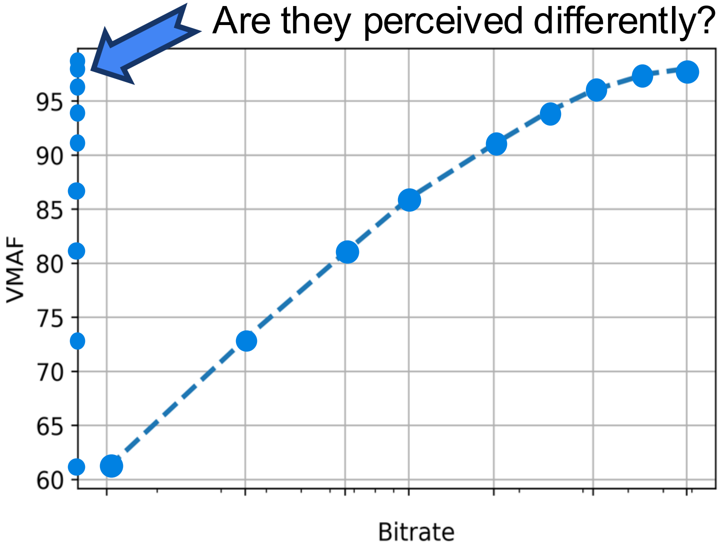
PyeongChang 2018

	Encoding Bitrate	Resolution
Rep. #1	18 Mbps	4K (60p)
Rep. #2	12.2 Mbps	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	2 Mbps	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	300 Kbps	320 x 180
Rep. #10	200 Kbps	320 x 180

Bitrate Ladder Construction

Quality-aware bitrate ladder construction





PyeongChang 2018

	Encoding Bitrate	Resolution
Rep. #1	18 Mbps	4K (60p)
Rep. #2	12.2 Mbps	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	2 Mbps	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	300 Kbps	320 x 180
Rep. #10	200 Kbps	320 x 180

Bitrate Ladder Construction

Original Video

18Mpbs

100% of people perceive them as similar

Bitrate Ladder Construction

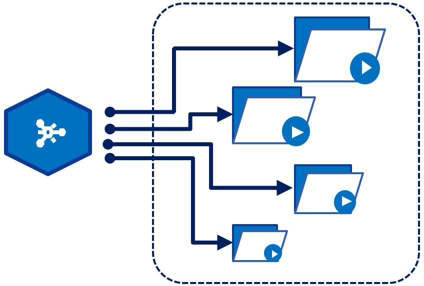
Original Video

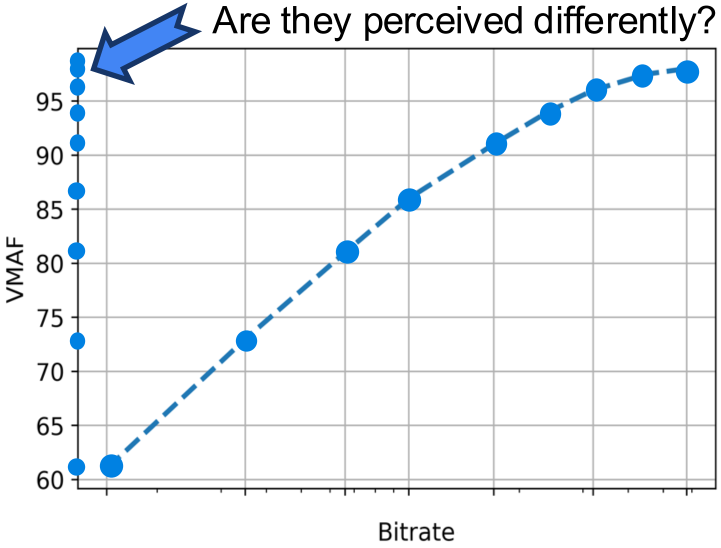
12Mpbs

98% of people perceive them as similar

Bitrate Ladder Construction

Quality-aware bitrate ladder construction



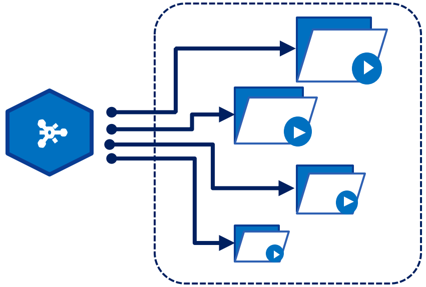


PyeongChang 2018

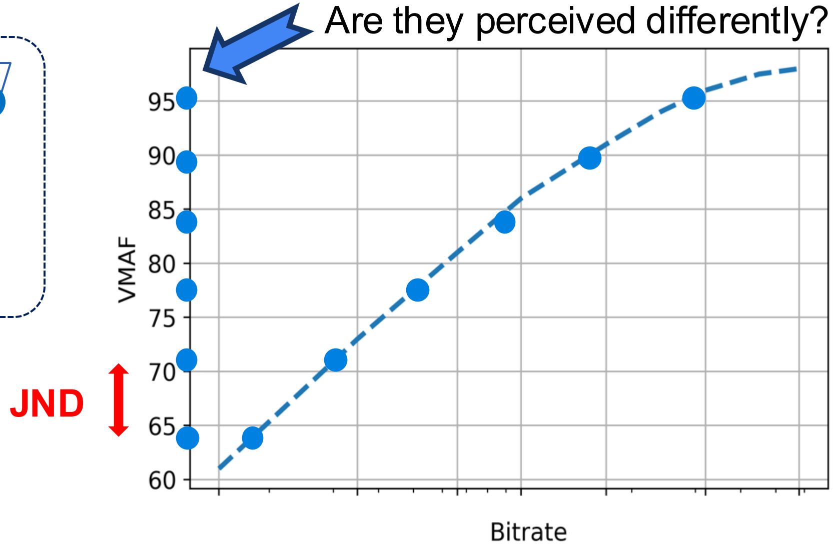
	Encoding Bitrate	Resolution
Rep. #1	18 Mbps	4K (60p)
Rep. #2	12.2 Mbps	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	2 Mbps	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	300 Kbps	320 x 180
Rep. #10	200 Kbps	320 x 180

Bitrate Ladder Construction

Quality-aware bitrate ladder construction



The minimum visual difference that can be perceived by HVS, *i.e.*, the difference between two adjacent perceptual distortion levels, refers as to one **Just Noticeable Difference** (JND)

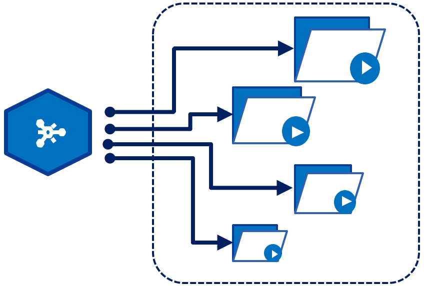


PyeongChang 2018

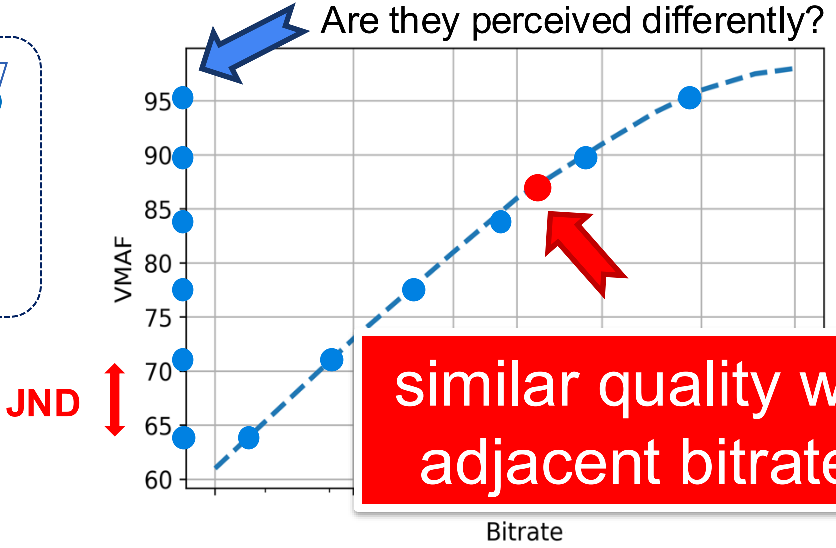
	Encoding Bitrate	Resolution
Rep. #1	18 Mbps	4K (60p)
Rep. #2	12.2 Mbps	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	2 Mbps	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	300 Kbps	320 x 180
Rep. #10	200 Kbps	320 x 180

Bitrate Ladder Construction

Quality-aware bitrate ladder construction



The minimum visual difference that can be perceived by HVS, *i.e.*, the difference between two adjacent perceptual distortion levels, refers as to one **Just Noticeable Difference** (JND)



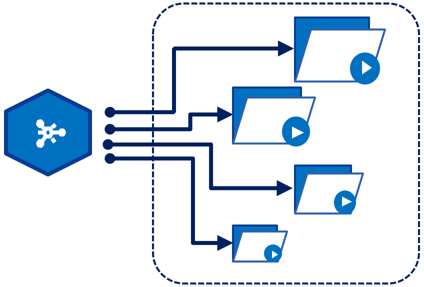
similar quality with adjacent bitrates

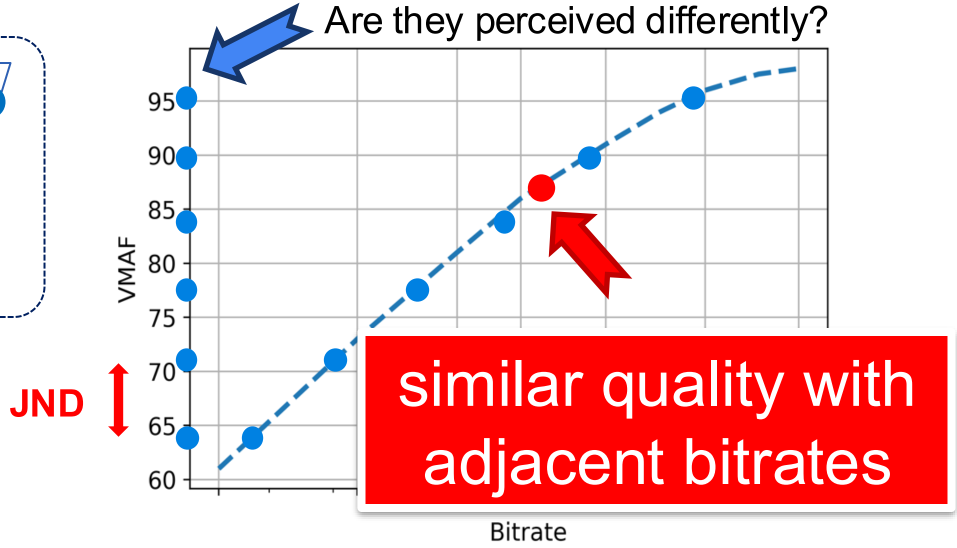
PyeongChang 2018

	Encoding Bitrate	Resolution
Rep. #1	~~18[illegible]bps~~	4K (60p)
Rep. #2	~~12.[illegible]bps~~	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	~~2[illegible]ps~~	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	~~30[illegible]bps~~	320 x 180
Rep. #10	200 Kbps	320 x 180

Bitrate Ladder Construction

Quality-aware bitrate ladder construction





The minimum visual difference that can be perceived by HVS, *i.e.*, the difference between two adjacent perceptual distortion levels, refers as to one **Just Noticeable Difference** (JND)

PyeongChang 2018

	Encoding Bitrate	Resolution
Rep. #1	~~18 bps~~	4K (60p)
Rep. #2	~~12. bps~~	2560x1440 (60p)
Rep. #3	4.7 Mbps	2K (60p)
Rep. #4	3.5 Mbps	1280x720 (60p)
Rep. #5	~~2 ps~~	1280 x 720
Rep. #6	1.2 Mbps	768 x 432
Rep. #7	750 Kbps	640 x 360
Rep. #8	500 Kbps	512 x 288
Rep. #9	~~30 bps~~	320 x 180
Rep. #10	200 Kbps	320 x 180

30% Bitrate Saving

[Menon, **Amirpour**, et al, 2022, IEEE ICME]

[Menon, **Amirpour**, et al, 2022, IEEE ICME]

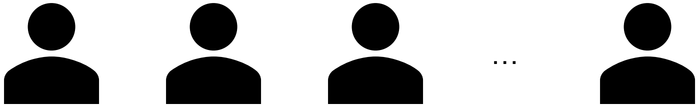
Just Noticeable Difference (JND)

SRC

QP=0

...

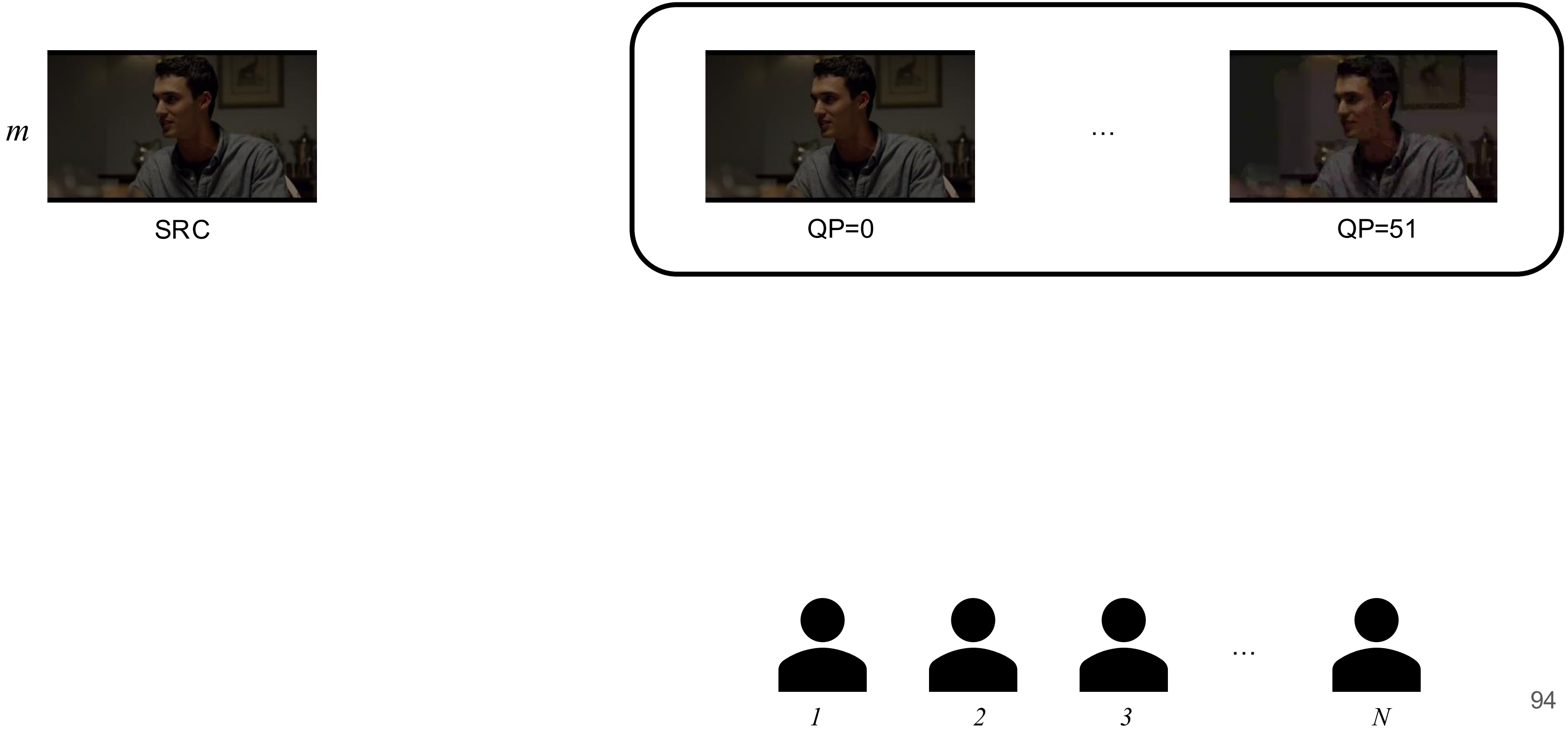
QP=51



...

Just Noticeable Difference (JND)

Just Noticeable Difference (JND)



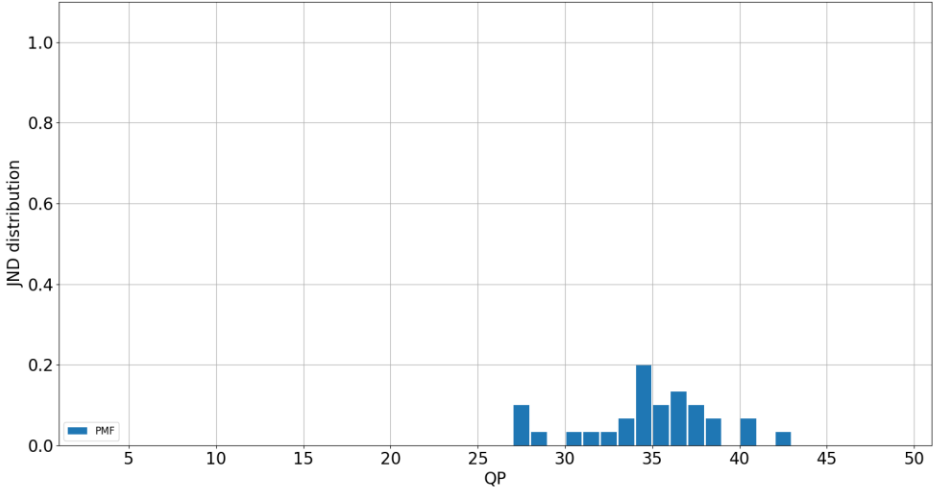
Just Noticeable Difference (JND)

m

SRC

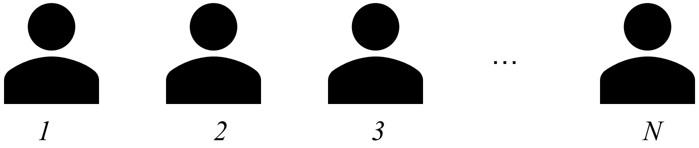
For video content m, VW-JND annotations for N reliable subjects :

$$J^m = \left[j_1^m, j_2^m, \dots, j_N^m \right]$$



Probability Mass Function (PMF) of J^m :

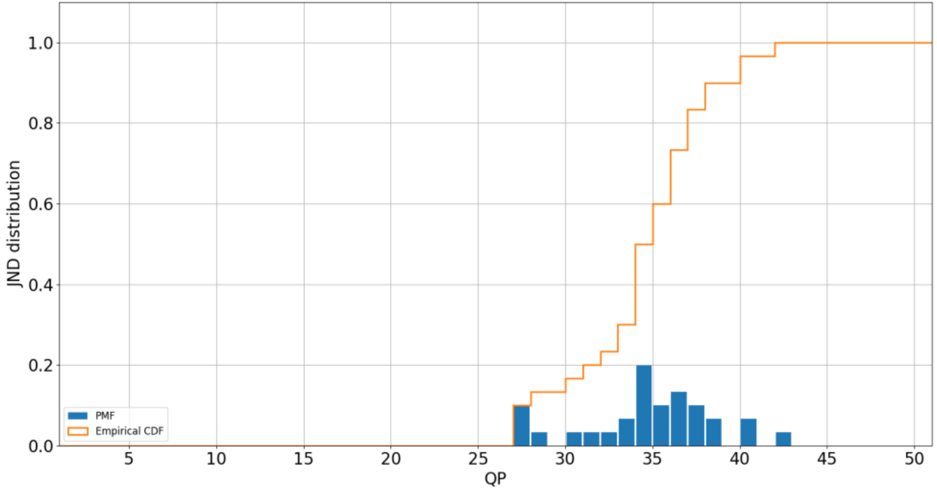
$$p^m(x) = P(\text{VW-JND} = x) = \frac{1}{N} \sum_{i=1}^{N} 1\left(j_i^m = x \right)$$



Just Noticeable Difference (JND)

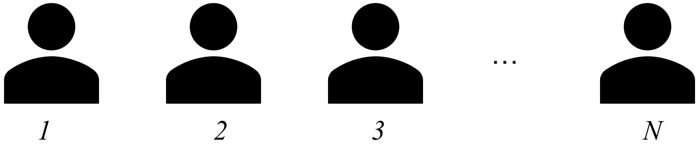
m

SRC



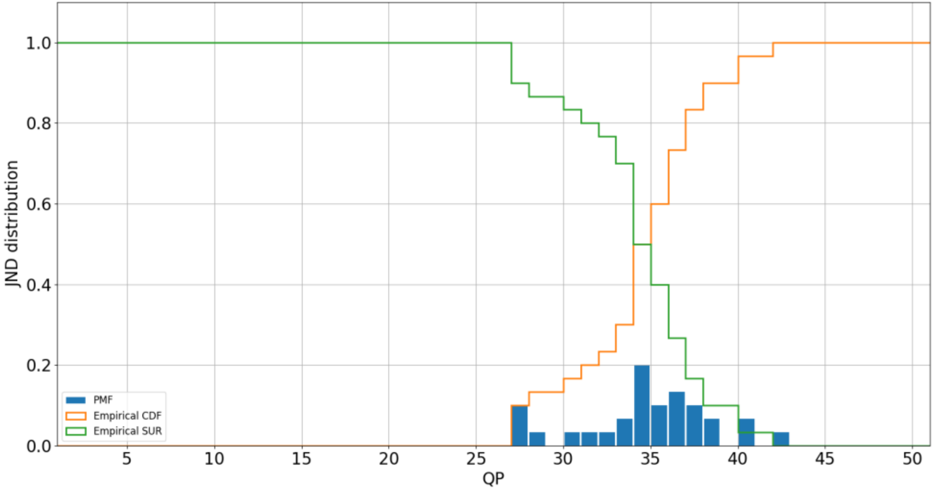
Cumulative Distribution Function (CDF) can be calculated from the PMF:

$$\mathrm{CDF}_{\mathrm{emp}}^{\mathrm{m}}(x) = P(\mathrm{VW\text{-}JND} \leq x) = \sum_{\omega < x} p^m(\omega)$$



Just Noticeable Difference (JND)

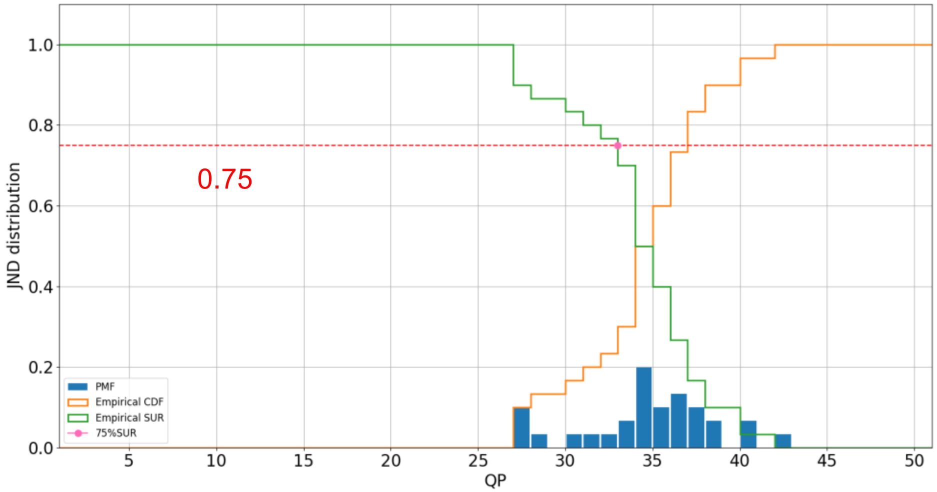
SRC



$$\mathrm{SUR}_{\mathrm{emp}}^{m}(x) = 1 - \mathrm{CDF}_{emp}^{m}(x)$$

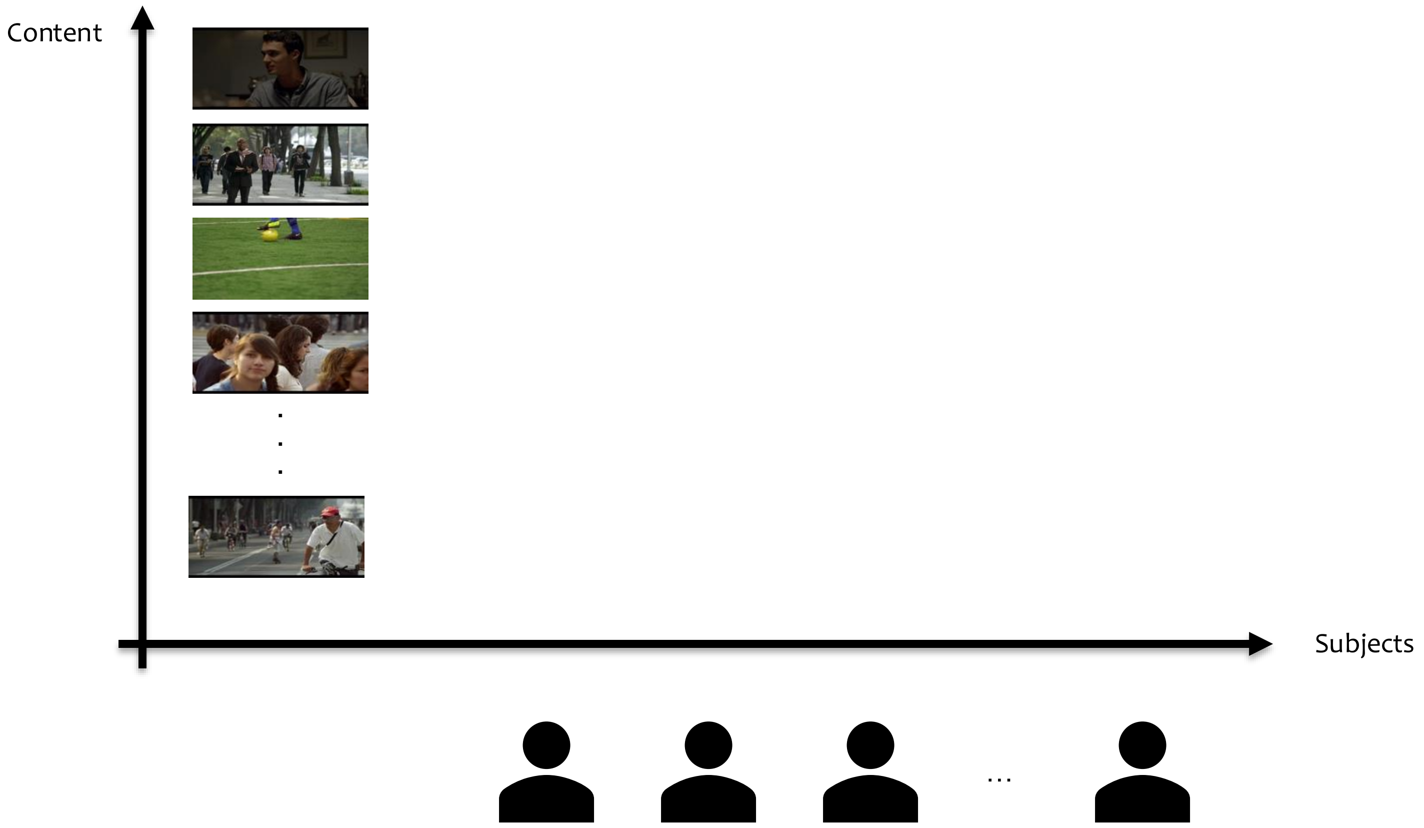
Just Noticeable Difference (JND)

SRC

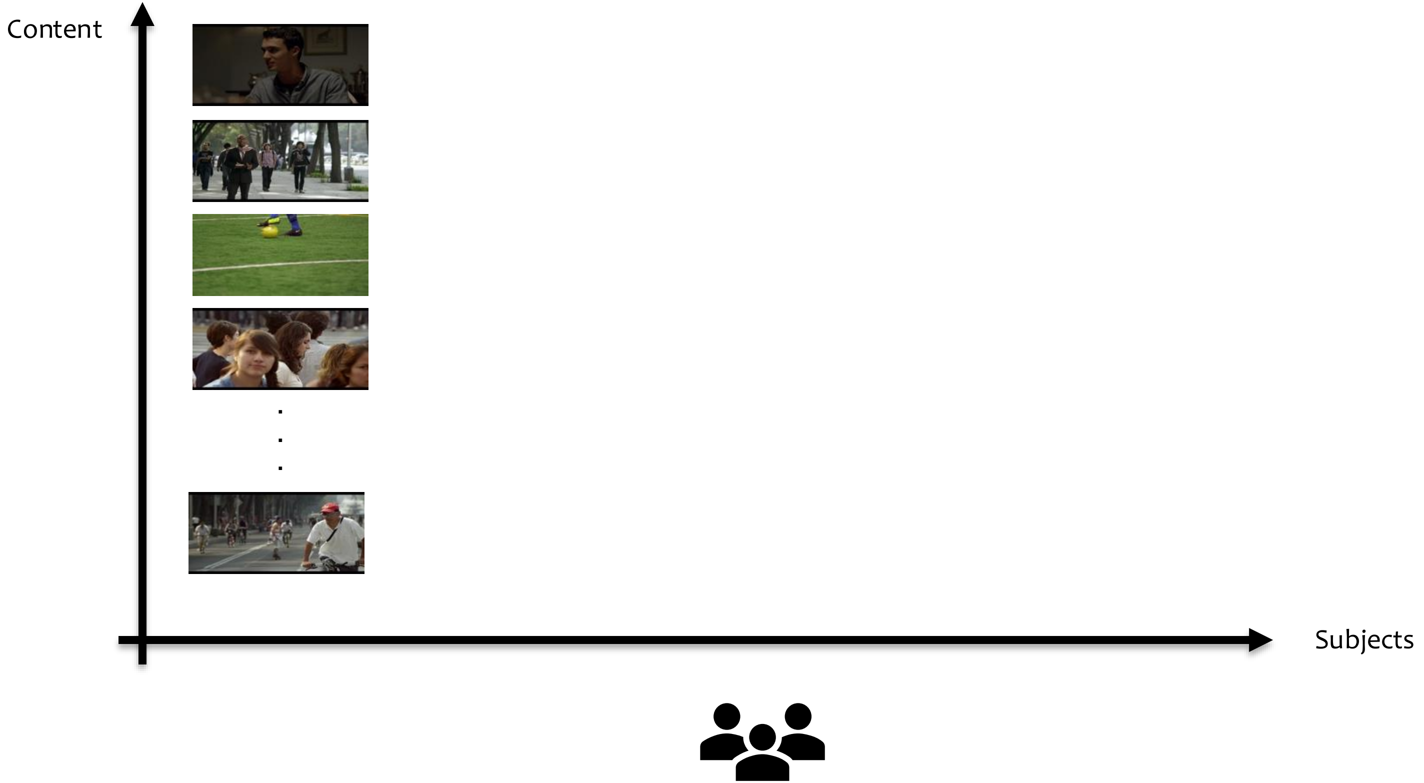


- Satisfied: didn’t perceive difference
- Not satisfied: perceived a difference

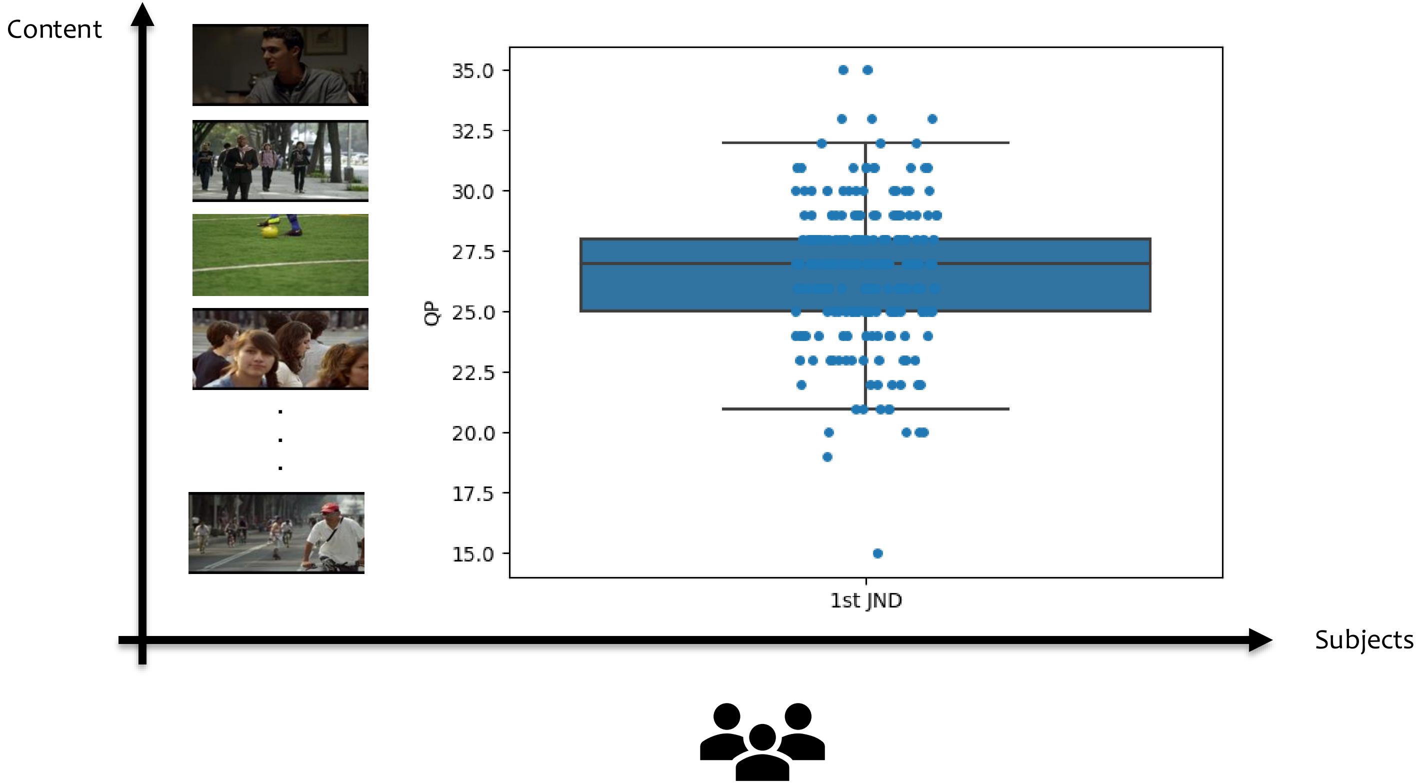
Just Noticeable Difference (JND)



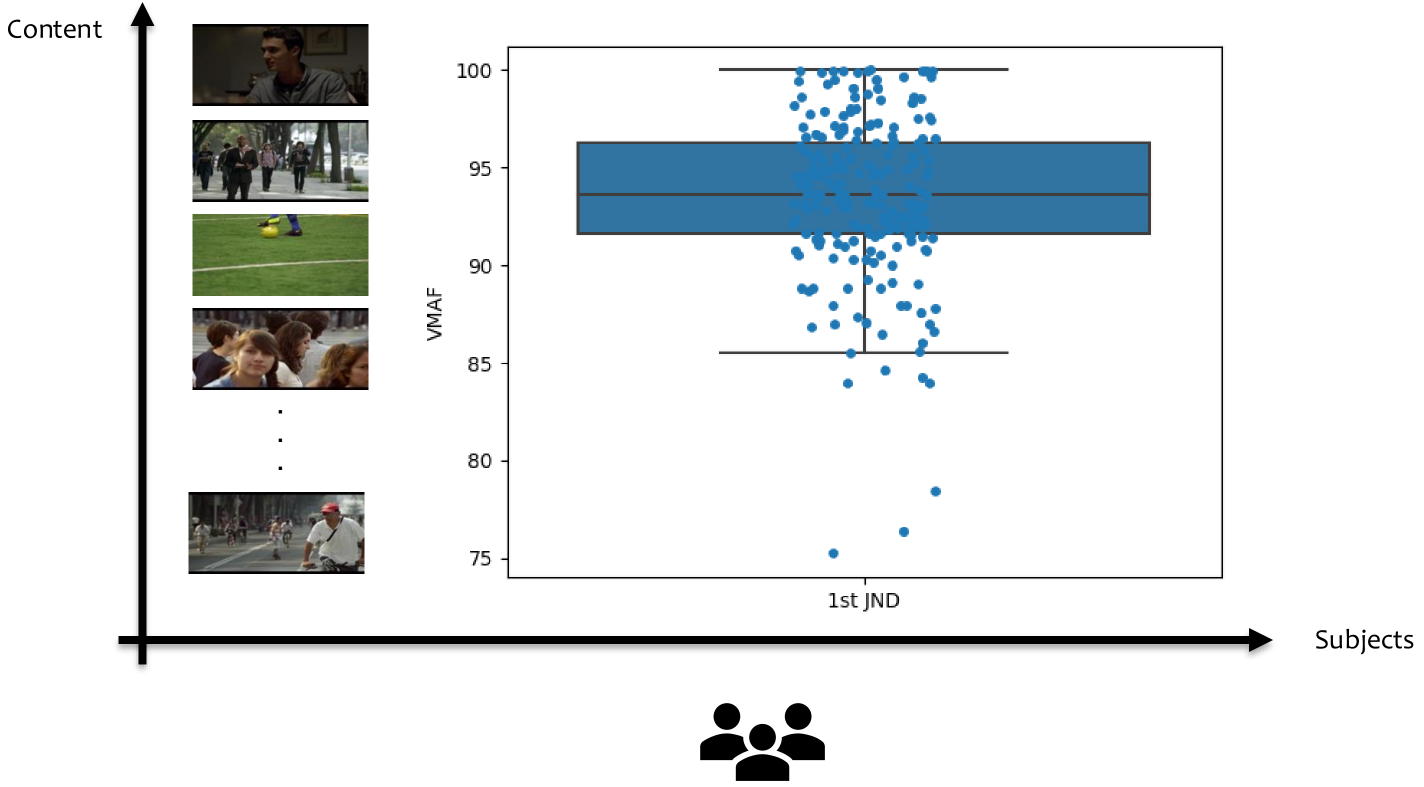
Just Noticeable Difference (JND)



Just Noticeable Difference (JND)



Just Noticeable Difference (JND)

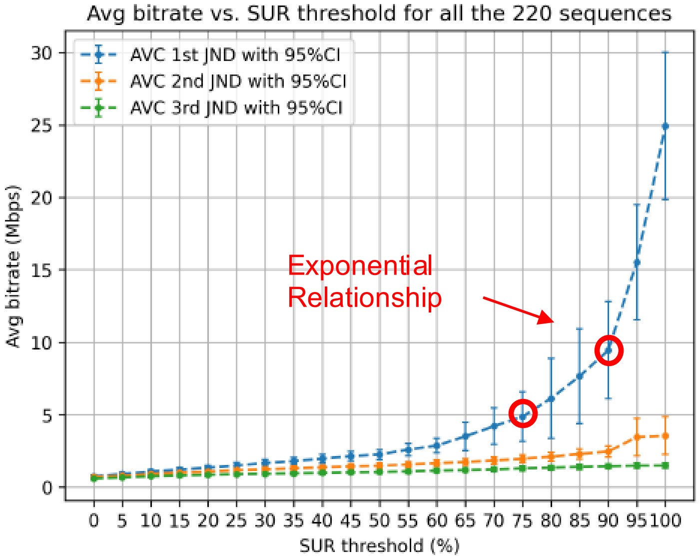


Uncertainty

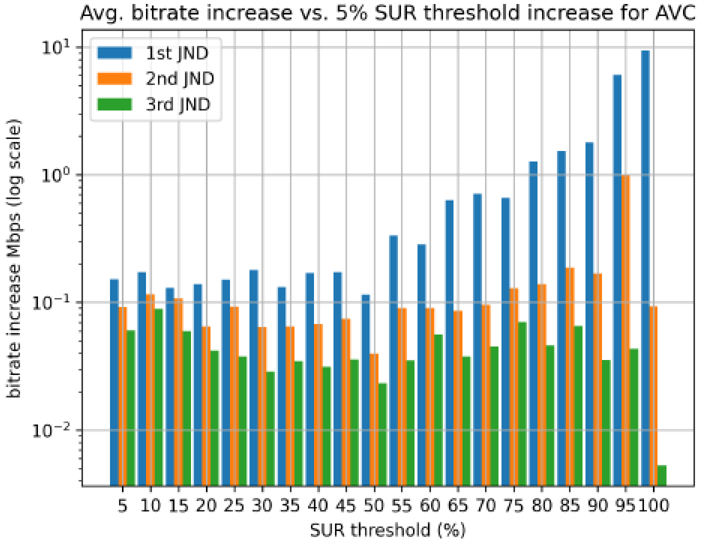
QP =33

QP =30

Bitrate as a function of SUR



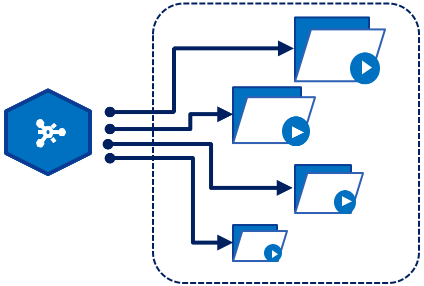
(a)



(b)

Bitrate Ladder Construction

Quality-aware bitrate ladder construction



JND for VMAF

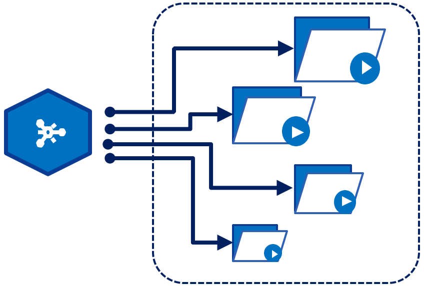
Netflix: 6

RheinMain Univ: 2

[**Amirpour**, et al, 2022, QoMEX]

Bitrate Ladder Construction

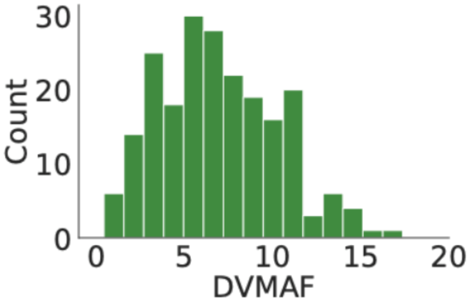
Quality-aware bitrate ladder construction



JND for VMAF

Netflix: 6

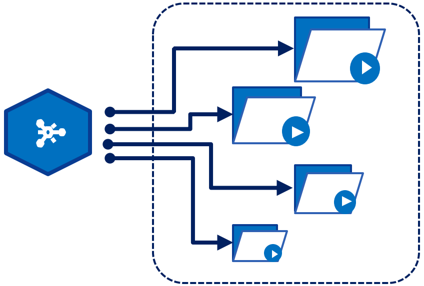
RheinMain Univ: 2



[**Amirpour**, et al, 2022, QoMEX]

Bitrate Ladder Construction

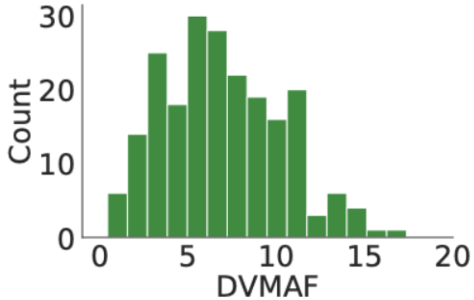
Quality-aware bitrate ladder construction

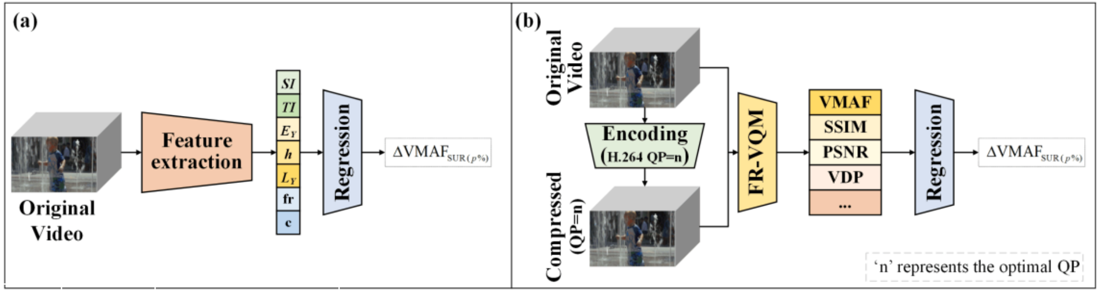


JND for VMAF

Netflix: 6

RheinMain: 2



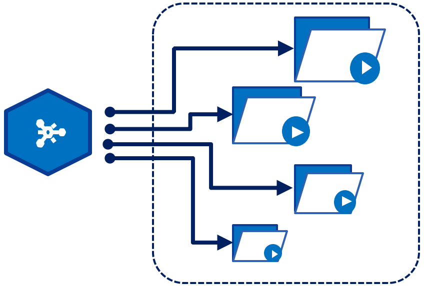


[**Amirpour**, et al, 2022, QoMEX]

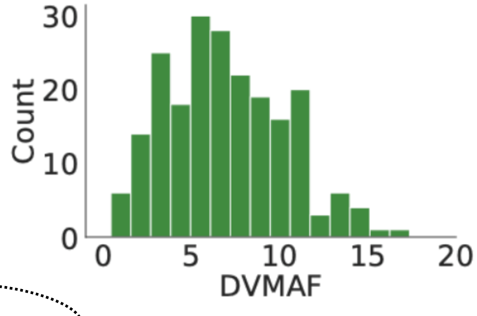
[Zhu, **Amirpour**, et al, 2023, IEEE VCIP]

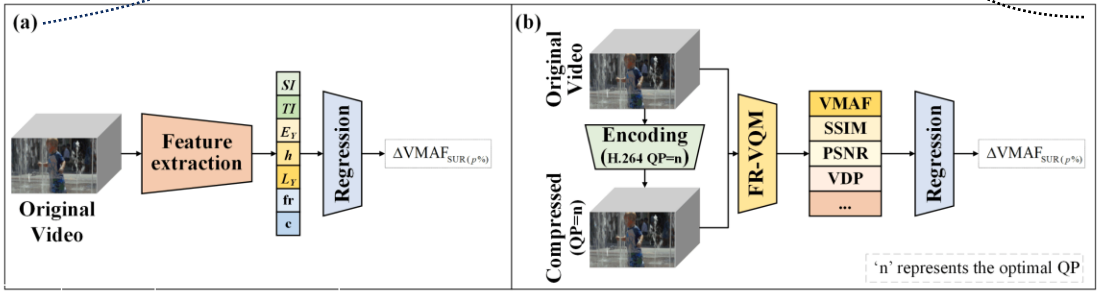
Bitrate Ladder Construction

Quality-aware bitrate ladder construction



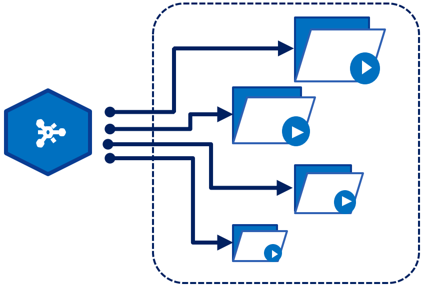
Model	Features	qp	MAE
Jan *et al.* [19]	$\Delta\mathrm{VMAF}_{\mathrm{SUR}(75\%)} = 2$	-	4.29
Kah *et al.* [20]	$\Delta\mathrm{VMAF}_{\mathrm{SUR}(75\%)} = 6$	-	2.59
Amirpour *et al.* [16]	SI, TI, E, h, L, c, fr	-	2.01
Ridge regression	vmaf, ssim, vdp, psnr, ms-ssim	30	1.77
Ridge regression	vmaf, ssim, vdp, psnr	30	1.77
Ridge regression	vmaf, ssim, vdp	29	1.78
Ridge regression	vmaf, vdp	29	1.78
Ridge regression	vmaf	29	1.80
XGBoost	vmaf, ssim, vdp, psnr, ms-ssim	28	1.73
XGBoost	vmaf, ssim, vdp, psnr	28	1.74
XGBoost	vmaf, ssim, vdp	28	1.73
XGBoost	vmaf, ssim	28	**1.67**
XGBoost	vmaf	28	1.72





[**Amirpour**, et al, 2022, QoMEX]
[Zhu, **Amirpour**, et al, 2023, IEEE VCIP]

Bitrate Ladder Construction

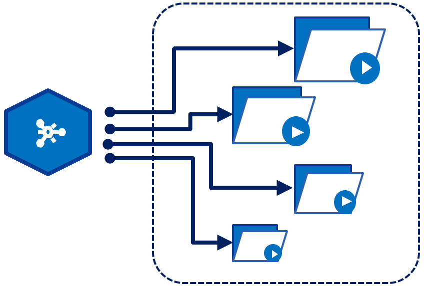


	Encoding Bitrate
Rep. #01	145 kbps
Rep. #02	300 kbps
Rep. #03	600 kbps
Rep. #04	900 kbps
Rep. #05	1600 kbps
Rep. #06	2400 kbps
Rep. #07	3400 kbps
Rep. #08	4500 kbps
Rep. #09	5800 kbps
Rep. #10	8100 kbps
Rep. #11	11600 kbps
Rep. #12	16800 kbps

Bitrates

- Quality-aware
- Network-aware

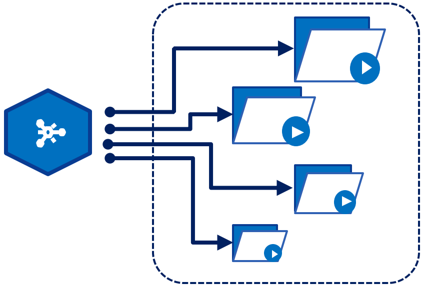
Bitrate Ladder Construction

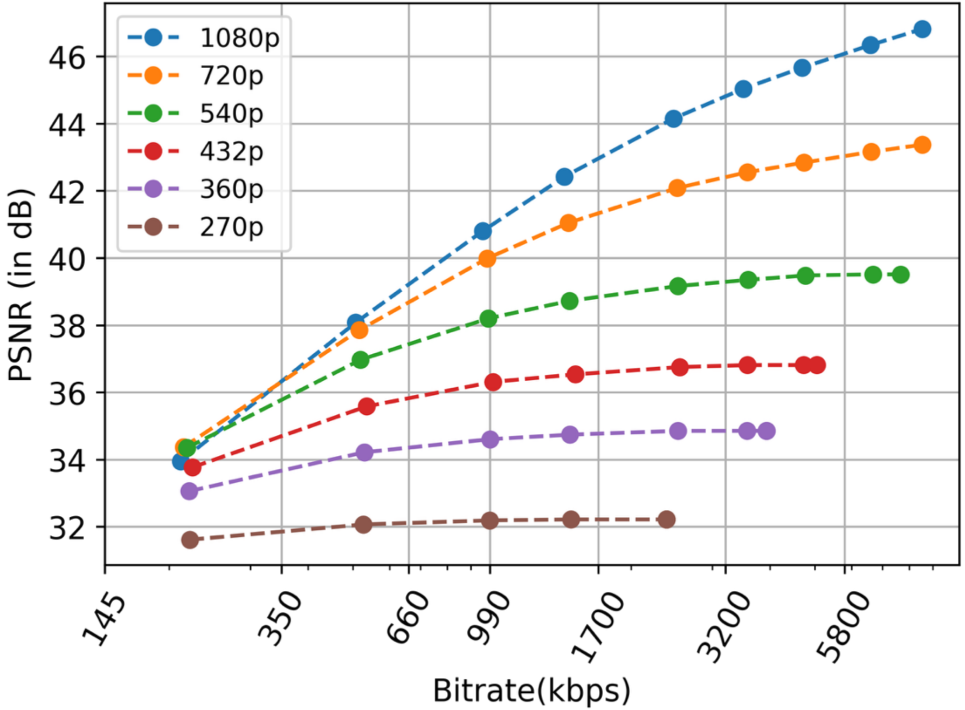


	Encoding Bitrate	Resolution
Rep. #01	145 kbps	640x360
Rep. #02	300 kbps	768x432
Rep. #03	600 kbps	960x540
Rep. #04	900 kbps	960x540
Rep. #05	1600 kbps	960x540
Rep. #06	2400 kbps	1280x720
Rep. #07	3400 kbps	2180x720
Rep. #08	4500 kbps	1920x1080
Rep. #09	5800 kbps	1920x1080
Rep. #10	8100 kbps	2560x1440
Rep. #11	11600 kbps	3840x2160
Rep. #12	16800 kbps	3840x2160

Bitrate Ladder Construction

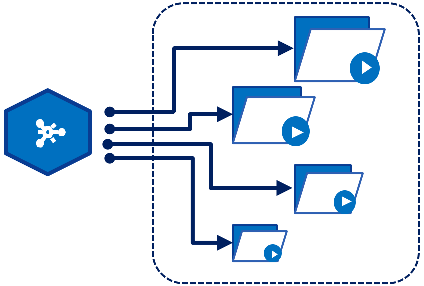
Per-title Encoding

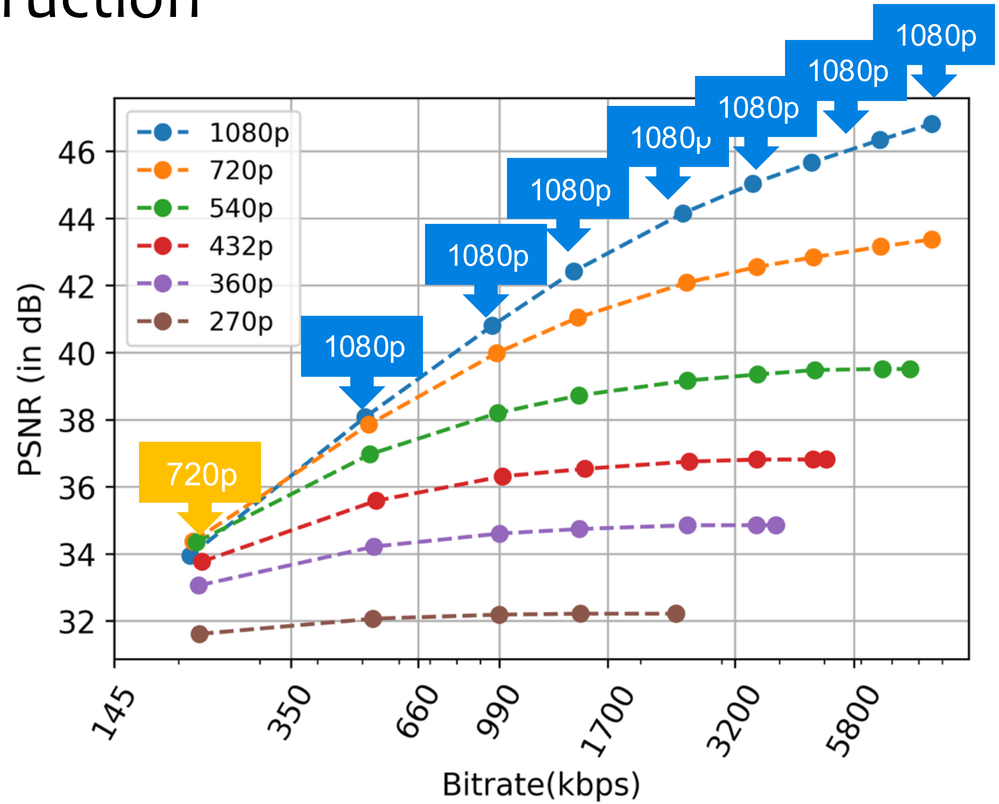




Bitrate Ladder Construction

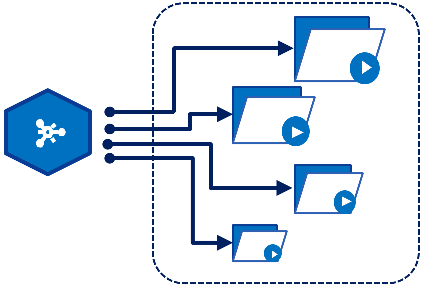
Per-title Encoding

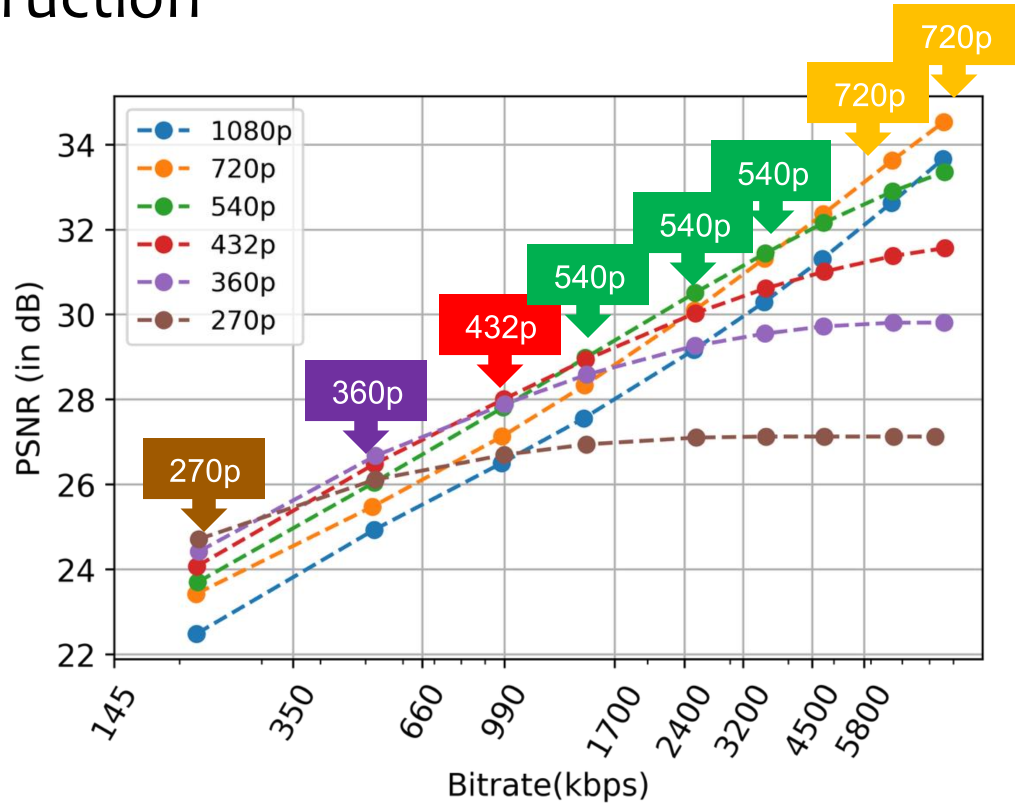




Bitrate Ladder Construction

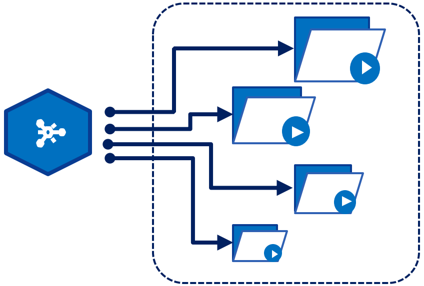
Per-title Encoding

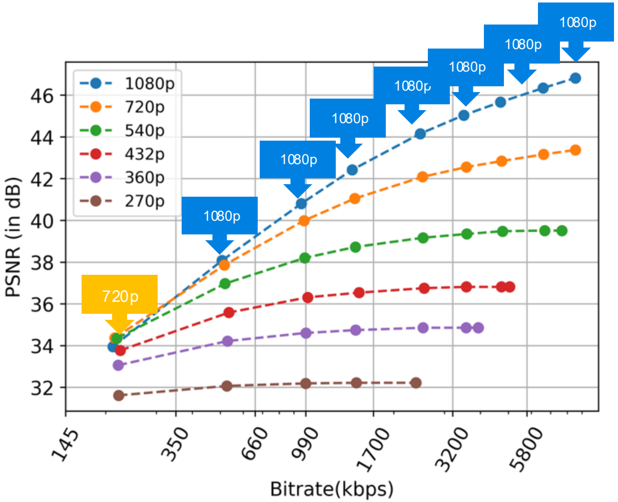


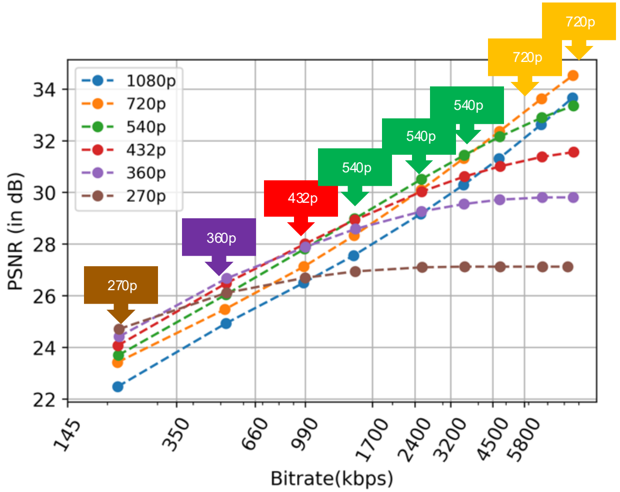


Bitrate Ladder Construction

Per-title Encoding





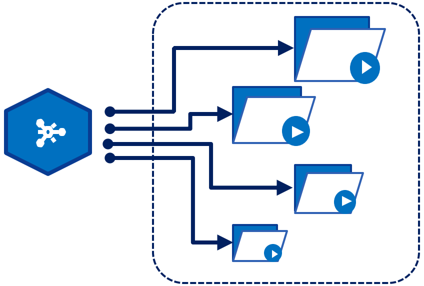


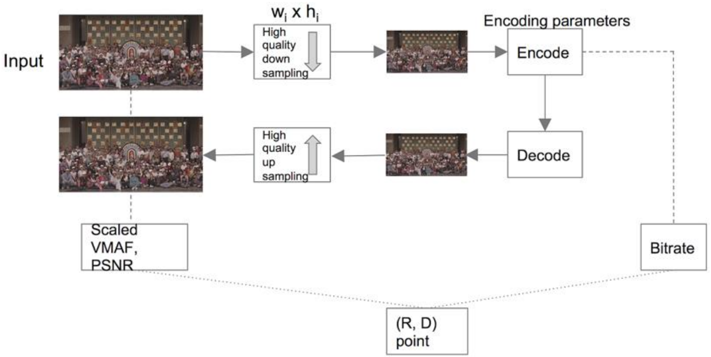
[De Cock, et al, 2016, IEEE ICME]

Bitrate Ladder Construction

Per-title Encoding

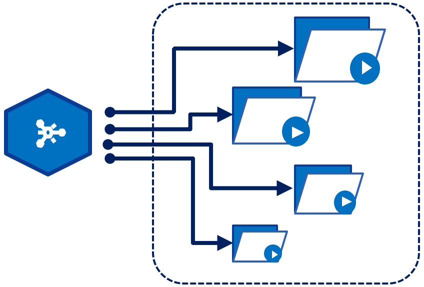
$$D_{Inverse}(VMAF) = \frac{1}{1+VMAF}$$

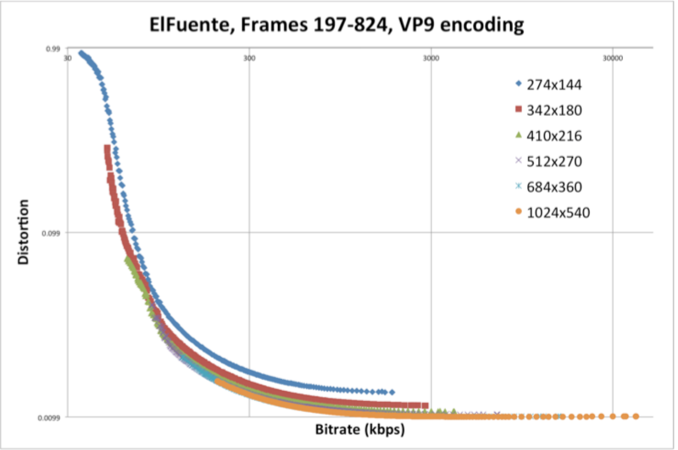


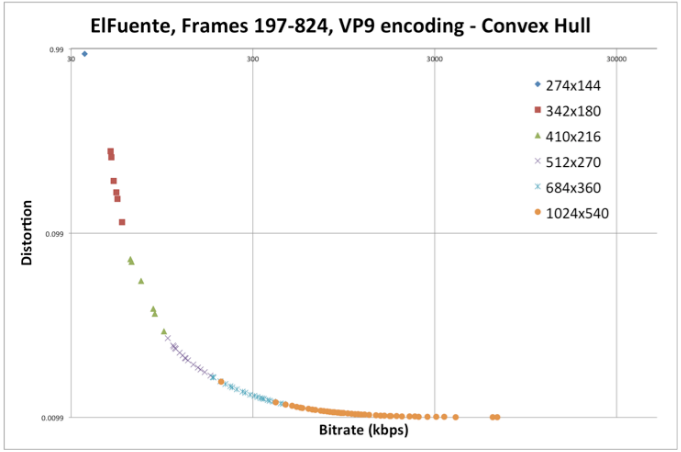


Bitrate Ladder Construction

Per-title Encoding

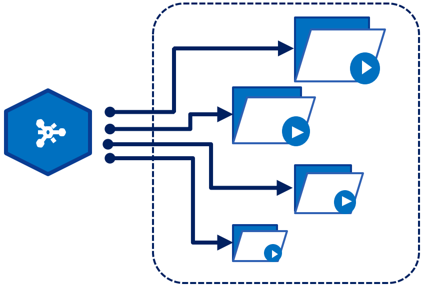


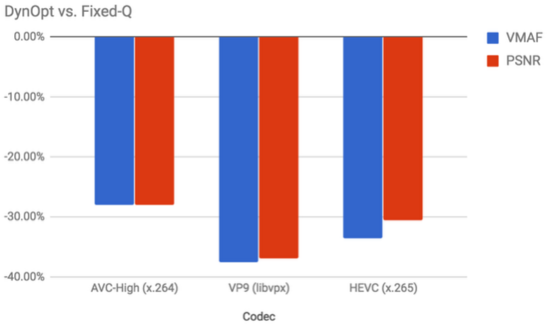


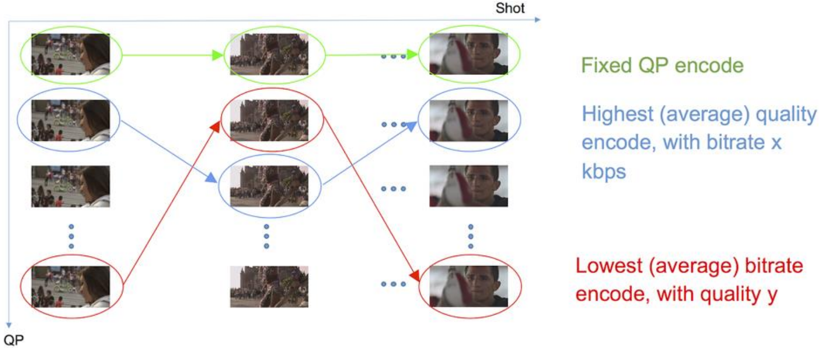


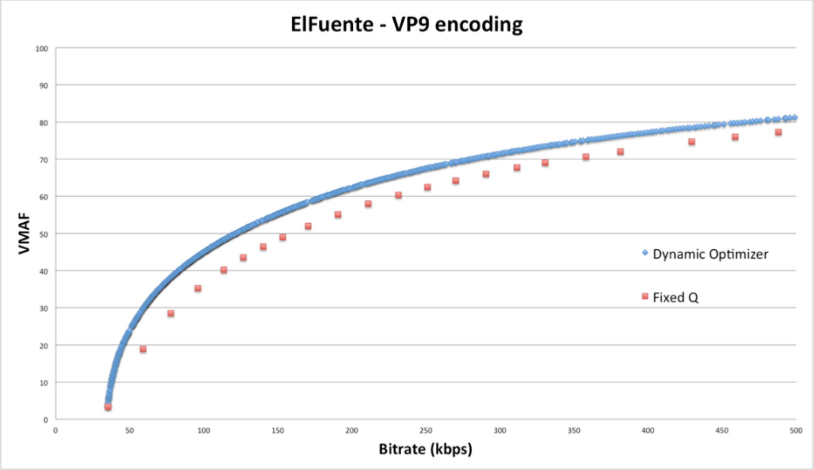
Bitrate Ladder Construction

Per-title Encoding





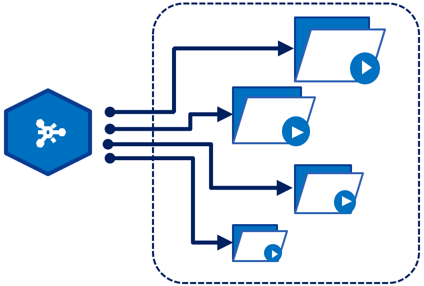




Universität Klagenfurt

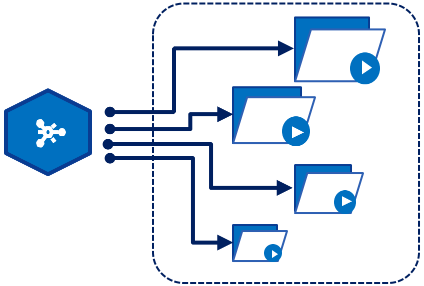
ATHENA
Christian Doppler (CD) Laboratory

Bitrate Ladder Construction



	Encoding Bitrate	Resolution
Rep. #01	145 kbps	640x360
Rep. #02	300 kbps	768x432
Rep. #03	600 kbps	960x540
Rep. #04	900 kbps	960x540
Rep. #05	1600 kbps	960x540
Rep. #06	2400 kbps	1280x720
Rep. #07	3400 kbps	2180x720
Rep. #08	4500 kbps	1920x1080
Rep. #09	5800 kbps	1920x1080
Rep. #10	8100 kbps	2560x1440
Rep. #11	11600 kbps	3840x2160
Rep. #12	16800 kbps	3840x2160

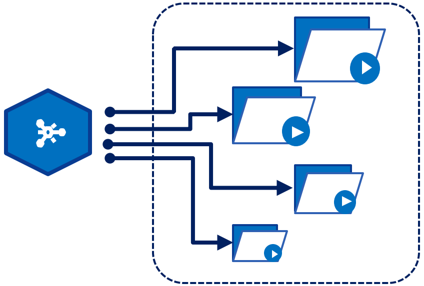
Bitrate Ladder Construction

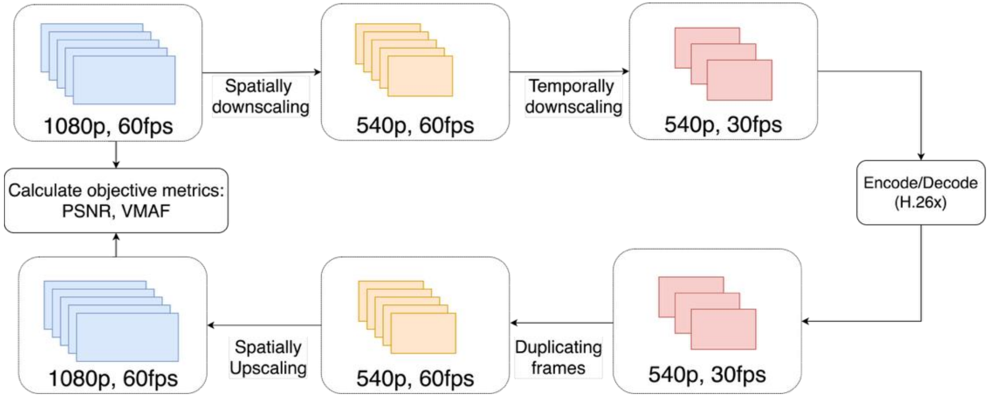


	Encoding Bitrate	Resolution	FR
Rep. #01	145 kbps	640x360	60
Rep. #02	300 kbps	768x432	60
Rep. #03	600 kbps	960x540	60
Rep. #04	900 kbps	960x540	60
Rep. #05	1600 kbps	960x540	60
Rep. #06	2400 kbps	1280x720	60
Rep. #07	3400 kbps	2180x720	60
Rep. #08	4500 kbps	1920x1080	60
Rep. #09	5800 kbps	1920x1080	60
Rep. #10	8100 kbps	2560x1440	60
Rep. #11	11600 kbps	3840x2160	60
Rep. #12	16800 kbps	3840x2160	60

Bitrate Ladder Construction

Per-title Encoding using Spatio-Temporal Resolutions (PSTR)





Bitrate Savings:

Resolution: 16%

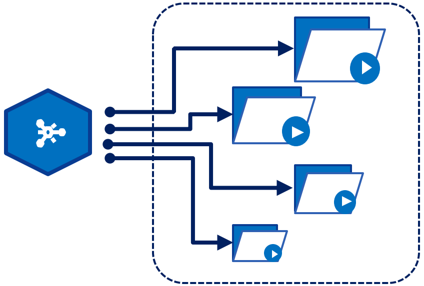
Resolution+Framerate: 33%

Table 1: Bitrate saving (BD-rate (%)) against encoding with 1080p, 120fps.

	Video characteristics				BD-Rate (PSNR)		BD-Rate (VMAF)	
	Source	SI	TI	framerate	state-of-the-art	PSTR	state-of-the-art	PSTR
Beauty	UVG	16.89	7.21	120	-26.38	-28.15	-32.05	-48.06
Bosphorus	UVG	29.76	4.70	120	-13.72	-17.06	-15.00	-29.41
Flowers	BVI-HFR	56.01	3.11	120	-11.85	-37.54	-9.72	-24.16
Golf	BVI-HFR	51.25	3.08	120	-17.01	-42.17	-8.92	-33.11
HoneyBee	UVG	25.04	2.38	120	-3.09	-15.38	-9.31	-29.28
Jockey	UVG	25.71	20.95	120	-34.36	-34.71	-37.74	-37.74
Pond	BVI-HFR	70.03	1.65	120	-12.60	-45.90	-6.34	-38.80
Typing	BVI-HFR	29.28	3.75	120	-19.87	-35.35	-17.16	-22.25
YachtRide	UVG	47.83	12.02	120	-11.25	-11.74	-14.56	-35.84
average					-16.68	-29.82	-16.75	-33.18

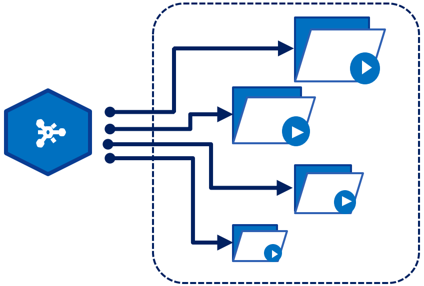
[**Amirpour**, et al, 2022, IEEE ICME]

Bitrate Ladder Construction



	Encoding Bitrate	Resolution	FR
Rep. #01	145 kbps	640x360	60
Rep. #02	300 kbps	768x432	60
Rep. #03	600 kbps	960x540	60
Rep. #04	900 kbps	960x540	50
Rep. #05	1600 kbps	960x540	50
Rep. #06	2400 kbps	1280x720	50
Rep. #07	3400 kbps	2180x720	30
Rep. #08	4500 kbps	1920x1080	30
Rep. #09	5800 kbps	1920x1080	30
Rep. #10	8100 kbps	2560x1440	30
Rep. #11	11600 kbps	3840x2160	24
Rep. #12	16800 kbps	3840x2160	24

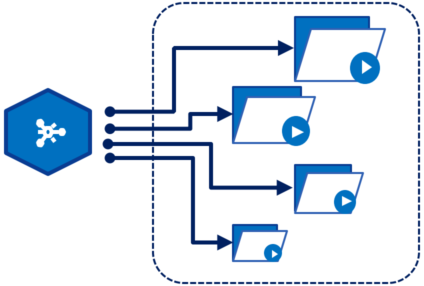
Bitrate Ladder Construction

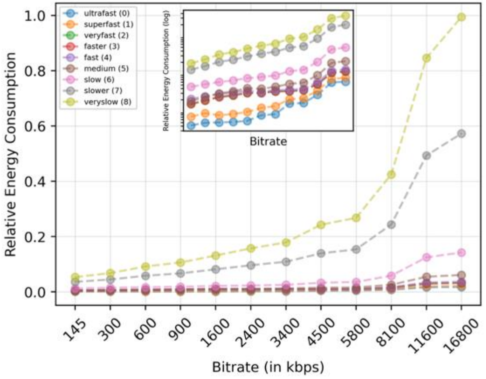


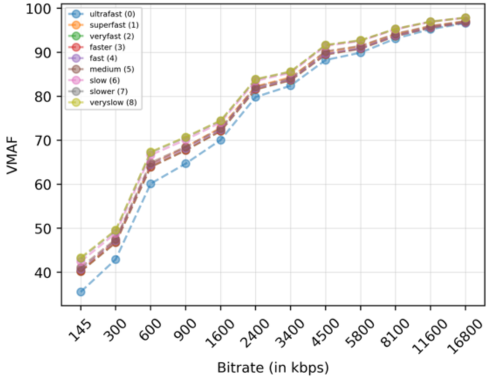
	Encoding Bitrate	Resolution	FR	Preset
Rep. #01	145 kbps	640x360	60	veryslow
Rep. #02	300 kbps	768x432	60	veryslow
Rep. #03	600 kbps	960x540	60	veryslow
Rep. #04	900 kbps	960x540	50	veryslow
Rep. #05	1600 kbps	960x540	50	veryslow
Rep. #06	2400 kbps	1280x720	50	veryslow
Rep. #07	3400 kbps	2180x720	30	veryslow
Rep. #08	4500 kbps	1920x1080	30	veryslow
Rep. #09	5800 kbps	1920x1080	30	veryslow
Rep. #10	8100 kbps	2560x1440	30	veryslow
Rep. #11	11600 kbps	3840x2160	24	veryslow
Rep. #12	16800 kbps	3840x2160	24	veryslow

Bitrate Ladder Construction

Energy Consumption





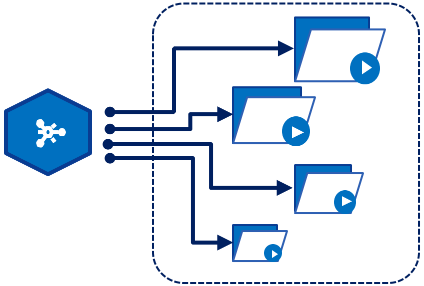


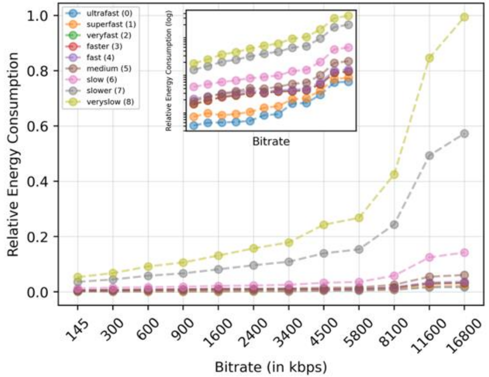
preset	name	ctu	min-cu-size	bframes	ref	me	merange	rc-lookahead
0	ultrafast	32	16	3	1	dia	57	5
1	superfast	32	8	3	1	hex	57	10
2	veryfast	64	8	4	2	hex	57	15
3	faster	64	8	4	2	hex	57	15
4	fast	64	8	4	3	hex	57	15
5	medium	64	8	4	3	hex	57	20
6	slow	64	8	4	4	star	57	25
7	slower	64	8	8	5	star	57	40
8	veryslow	64	8	8	5	star	57	40
9	placebo	64	8	8	5	star	92	60

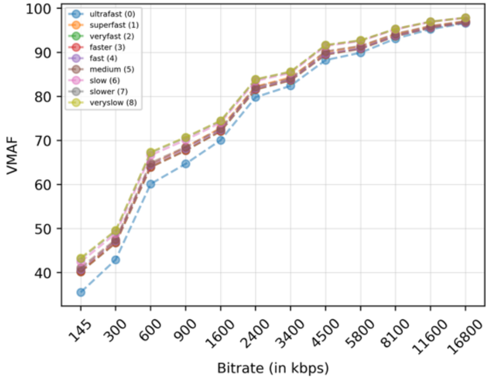
[**Amirpour**, et al, 2023, IEEE ICME]

Bitrate Ladder Construction

Energy Consumption



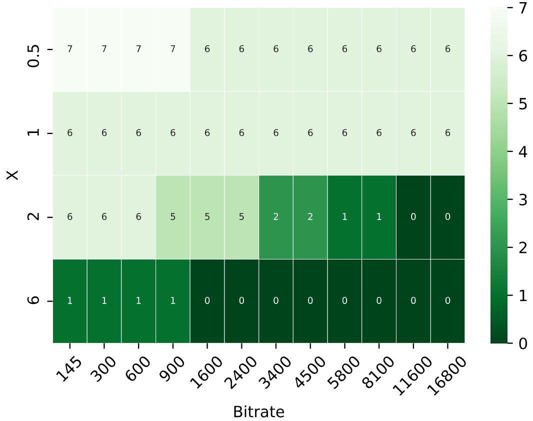




Preset Selection:

Energy Saving: 70%

Quality drop: 0.2 VMAF

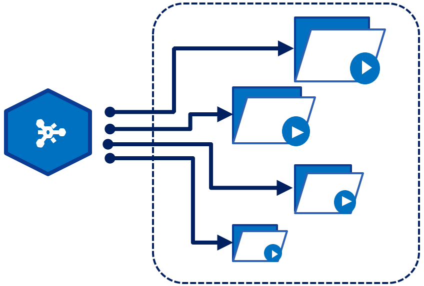


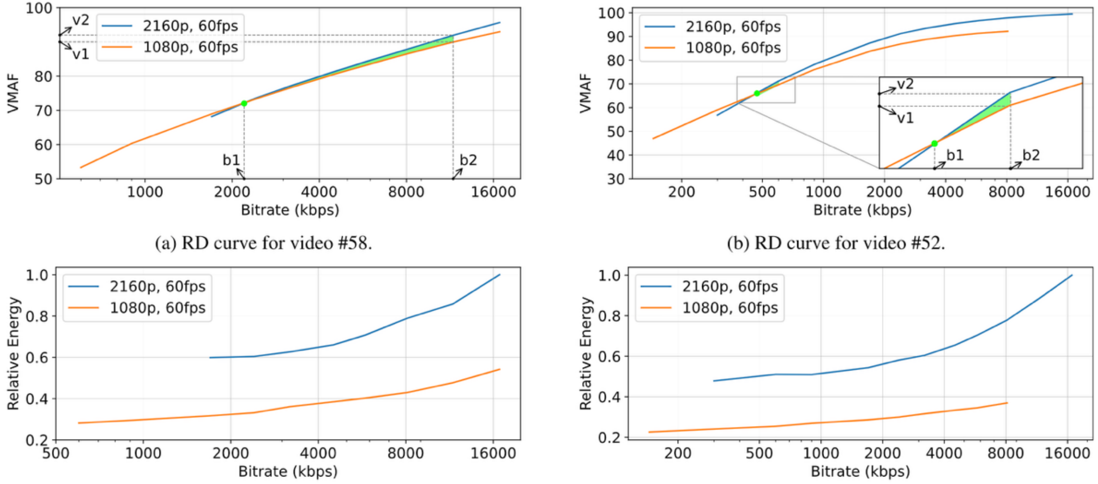
Model	X	0.5	1	2	6
avg	BD-VMAF	-0.22	-0.51	-1.53	-2.91
	$E_{saving}(\%)$	68.33	83.49	90.94	97.75
RF	BD-VMAF	-0.15	-0.39	-0.86	-1.92
	$E_{saving}(\%)$	71.08	82.35	90.86	97.20

[**Amirpour**, et al, 2023, IEEE ICME]

Bitrate Ladder Construction

Energy Consumption





(a) RD curve for video #58.

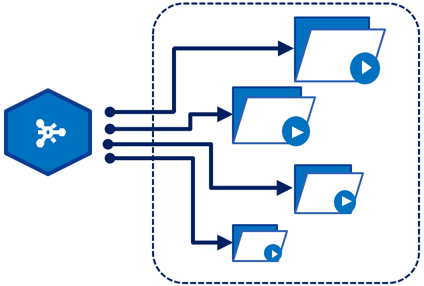
(b) RD curve for video #52.

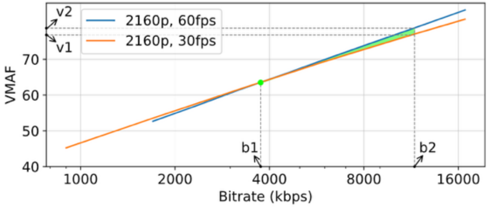
(c) relative decoding energy consumption for video #58.

(d) relative decoding energy consumption for video #52.

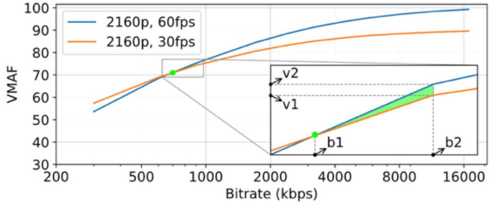
Bitrate Ladder Construction

Energy Consumption

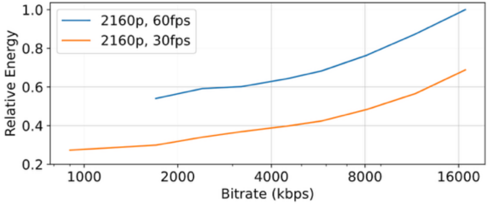




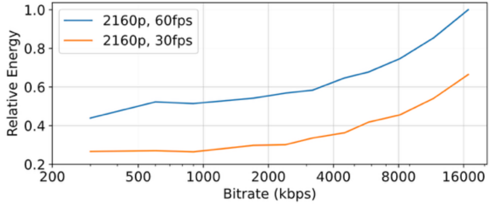
(a) RD curve for video #38.



(b) RD curve for video #65.



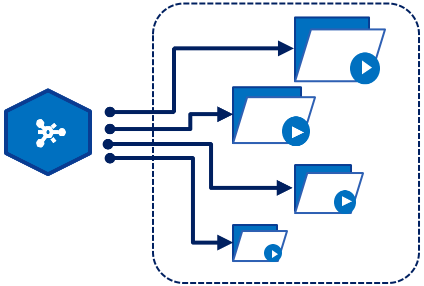
(c) relative decoding energy consumption for video #38.



(d) relative decoding energy consumption for video #65.

Bitrate Ladder Construction

Energy Consumption



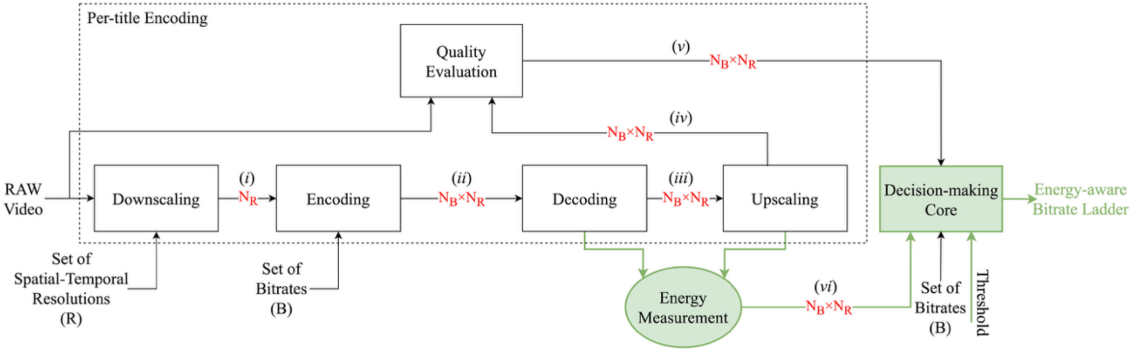
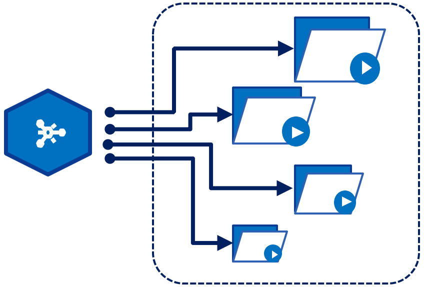
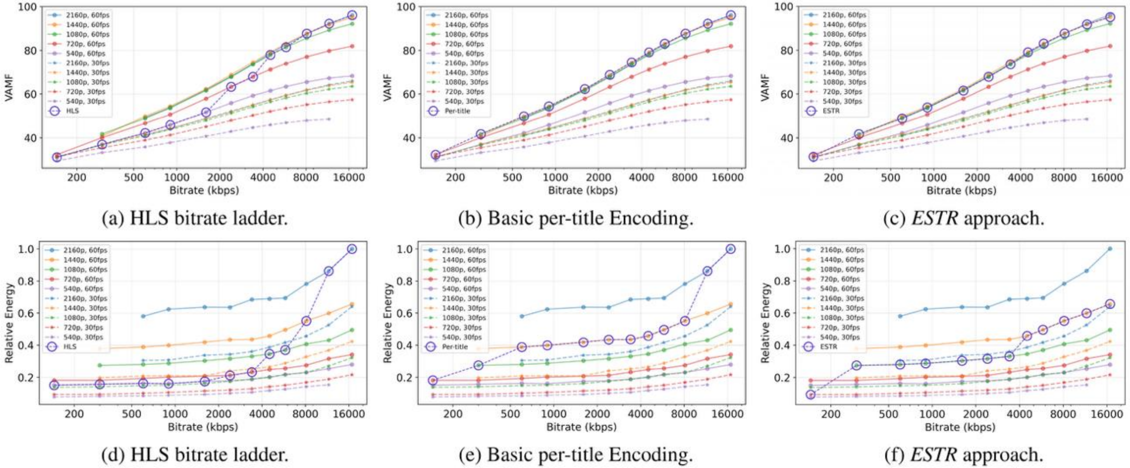


FIGURE 4. **The workflow of *ESTR*.**

Bitrate Ladder Construction

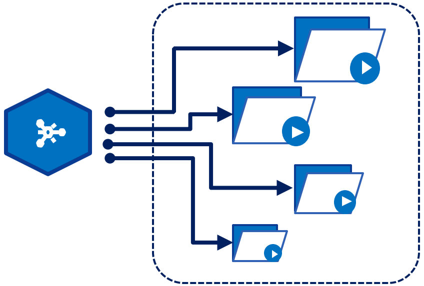
Energy Consumption





Bitrate Ladder Construction

Energy Consumption

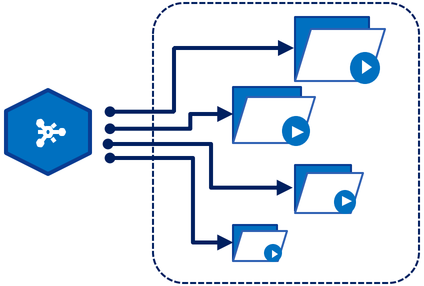


τ	BD-Rate %	BD-VMAF	BDDE %						
			Device 1 (Apple Mac)			Device 2 (Lenovo Linux)			Device 3 (Lenovo Windows)
			FFmpeg	HM	FFplay	FFmpeg	HM	FFplay	VLC (Hardware Accelerated)
0.50	-2.44	0.27	-21.37	-22.50	-19.58	-20.54	-21.56	-18.03	-22.35
0.75	-2.12	0.21	-24.94	-26.80	-23.79	-23.30	-25.76	-21.21	-27.14
1.00	-1.54	0.13	-27.69	-30.10	-27.04	-25.26	-29.01	-23.40	-31.93
1.25	-0.73	0.00	-31.18	-34.54	-31.15	-28.11	-33.27	-26.60	-36.75
1.50	0.09	-0.15	-33.84	-37.55	-34.25	-30.07	-36.20	-28.80	-40.00
1.75	1.09	-0.30	-36.44	-40.31	-37.15	-31.93	-38.90	-30.72	-43.35
2.00	2.52	-0.49	-39.20	-43.25	-40.22	-33.96	-41.86	-32.87	-46.37

TABLE 4. The performance of *ESTR* compared to the basic per-title encoding across various codecs.

τ	AVC (FFmpeg libx264)			HEVC (FFmpeg libx265)			VVC (VVdeC)		
	BD-Rate %	BD-VMAF	BDDE %	BD-Rate %	BD-VMAF	BDDE %	BD-Rate %	BD-VMAF	BDDE %
0.50	-5.01	0.83	-14.46	-2.44	0.27	-21.37	-0.70	0.08	-17.24
0.75	-4.83	0.79	-15.29	-2.12	0.21	-24.94	-0.01	0.00	-20.97
1.00	-4.46	0.74	-15.73	-1.54	0.13	-27.69	0.93	-0.11	-24.40
1.25	-4.17	0.64	-16.23	-0.73	0.00	-31.18	1.86	-0.23	-26.89
1.50	-4.02	0.57	-16.52	0.09	-0.15	-33.84	3.48	-0.39	-29.53
1.75	-3.65	0.47	-17.08	1.09	-0.30	-36.44	5.36	-0.56	-31.69
2.00	-2.99	0.37	-17.47	2.52	-0.49	-39.20	7.02	-0.73	-33.59

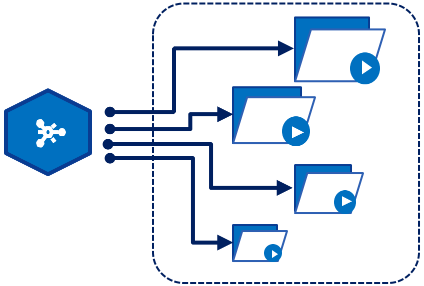
Bitrate Ladder Construction



	Encoding Bitrate	Resolution	FR	Preset	Bit depth
Rep. #01	145 kbps	640x360	60	veryslow	10
Rep. #02	300 kbps	768x432	60	veryslow	10
Rep. #03	600 kbps	960x540	60	slower	10
Rep. #04	900 kbps	960x540	60	slow	10
Rep. #05	1600 kbps	960x540	60	medium	10
Rep. #06	2400 kbps	1280x720	60	medium	10
Rep. #07	3400 kbps	2180x720	60	fast	10
Rep. #08	4500 kbps	1920x1080	60	faster	10
Rep. #09	5800 kbps	1920x1080	60	veryfast	10
Rep. #10	8100 kbps	2560x1440	60	veryfast	10
Rep. #11	11600 kbps	3840x2160	60	superfast	10
Rep. #12	16800 kbps	3840x2160	60	ultrafast	10

Bitrate Ladder Construction

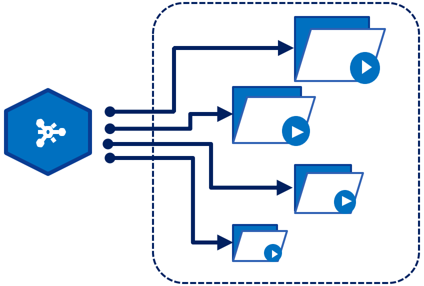
Energy Consumption



Factor	8-bit	10-bit
File size	Smaller	Slightly larger (5–20%)
Banding risk	High	Low
Color fidelity	Limited	Much better
Compatibility	Higher	Needs modern devices/codecs
Processing load	Lower	Slightly higher

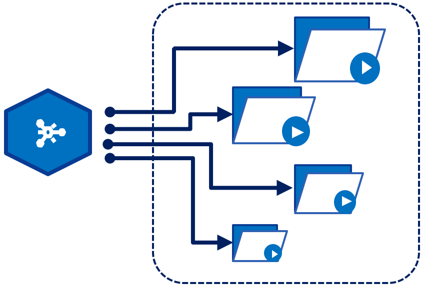
Bitrate Ladder Construction

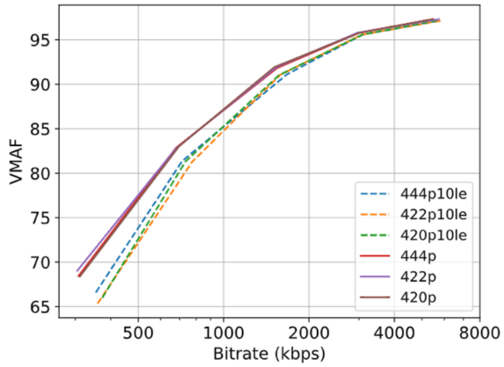
Energy Consumption



Bitrate Ladder Construction

Energy Consumption





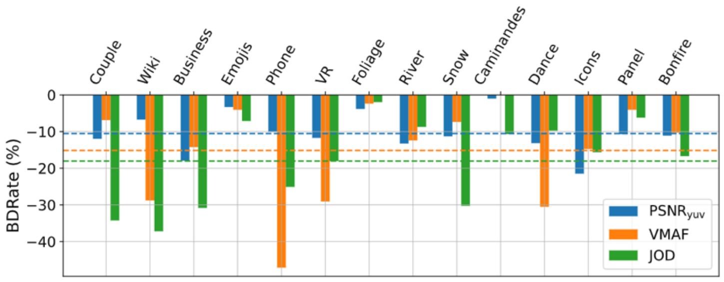


Fig. 4: BDR (%) for convex hull compared to YUV444p10le.

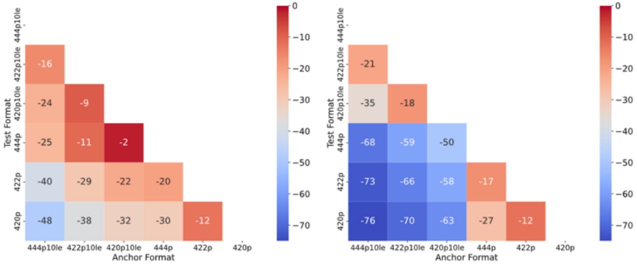
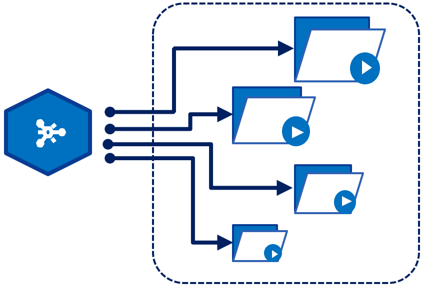


Fig. 6: (left) BDEE and (right) BDDE heatmaps.

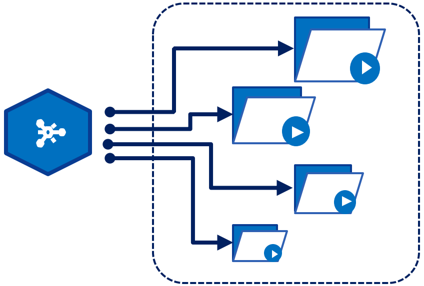
Bitrate Ladder Construction

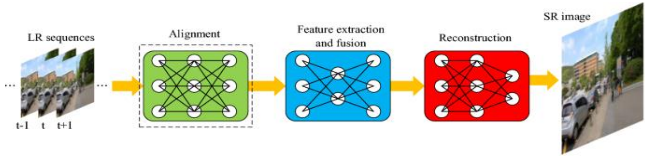


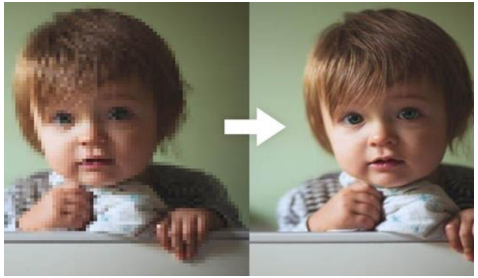
	Encoding Bitrate	Resolution	FR	Preset	Bit depth
Rep. #01	145 kbps	640x360	60	veryslow	8
Rep. #02	300 kbps	768x432	60	veryslow	8
Rep. #03	600 kbps	960x540	60	slower	9
Rep. #04	900 kbps	960x540	60	slow	9
Rep. #05	1600 kbps	960x540	60	medium	9
Rep. #06	2400 kbps	1280x720	60	medium	10
Rep. #07	3400 kbps	2180x720	60	fast	10
Rep. #08	4500 kbps	1920x1080	60	faster	10
Rep. #09	5800 kbps	1920x1080	60	veryfast	10
Rep. #10	8100 kbps	2560x1440	60	veryfast	10
Rep. #11	11600 kbps	3840x2160	60	superfast	10
Rep. #12	16800 kbps	3840x2160	60	ultrafast	10

Bitrate Ladder Construction

CPU vs GPU devices

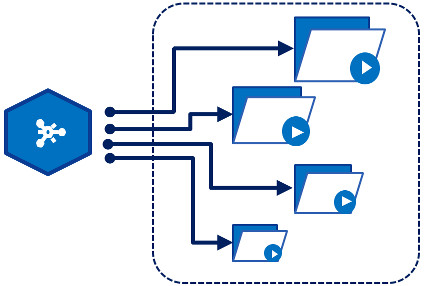






Bitrate Ladder Construction

CPU vs GPU devices

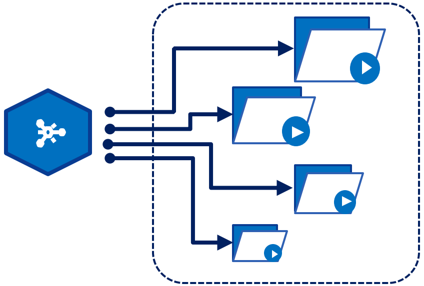


	Encoding Bitrate	Resolution
Rep.#01	145 kbps	640x360
Rep.#02	300 kbps	768x432
Rep.#03	600 kbps	960x540
Rep.#04	900 kbps	960x540
Rep.#05	1600 kbps	960x540
Rep.#06	2400 kbps	1280x720
Rep.#07	3400 kbps	1280x720
Rep.#08	4500 kbps	1920x1080
Rep.#09	5800 kbps	1920x1080
Rep.#10	8100 kbps	2560x1440
Rep.#11	11600 kbps	3840x2160
Rep.#12	16800 kbps	3840x2160

	Encoding Bitrate	Resolution
Rep.#01	100 kbps	480x270
Rep.#02	200 kbps	480x270
Rep.#03	400 kbps	640x360
Rep.#04	700 kbps	640x360
Rep.#05	1200 kbps	768x432
Rep.#06	1800 kbps	960x540
Rep.#07	2500 kbps	960x540
Rep.#08	3500 kbps	960x540
Rep.#09	4500 kbps	1280x720
Rep.#10	6400 kbps	1280x720
Rep.#11	9200 kbps	1280x720
Rep.#12	13000 kbps	1920x1080

Bitrate Ladder Construction

CPU vs GPU devices



	Encoding Bitrate	Resolution
Rep. #01	145 kbps	640x360
Rep. #02	300 kbps	768x432
Rep. #03	600 kbps	960x540
Rep. #04	900 kbps	960x540
Rep. #05	1600 kbps	960x540
Rep. #06	2400 kbps	1280x720
Rep. #07	3400 kbps	1280x720
Rep. #08	4500 kbps	1920x1080
Rep. #09	5800 kbps	1920x1080
Rep. #10	8100 kbps	2560x1440
Rep. #11	11600 kbps	3840x2160
Rep. #12	16800 kbps	3840x2160

	Encoding Bitrate	Resolution
Rep. #01	100 kbps	480x270
Rep. #02	200 kbps	480x270
Rep. #03	400 kbps	640x360
Rep. #04	700 kbps	640x360
Rep. #05	1200 kbps	768x432
Rep. #06	1800 kbps	960x540
Rep. #07	2500 kbps	960x540
Rep. #08	3500 kbps	960x540
Rep. #09	4500 kbps	1280x720
Rep. #10	6400 kbps	1280x720
Rep. #11	9200 kbps	1280x720
Rep. #12	13000 kbps	1920x1080

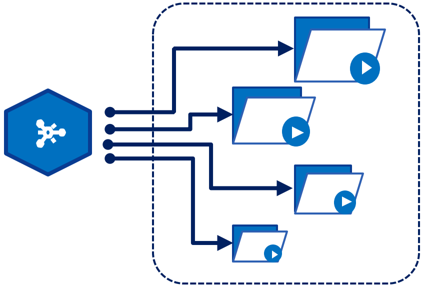
Challenges

- Backward Compatibility
- Increasing cost
- DNN Reliability

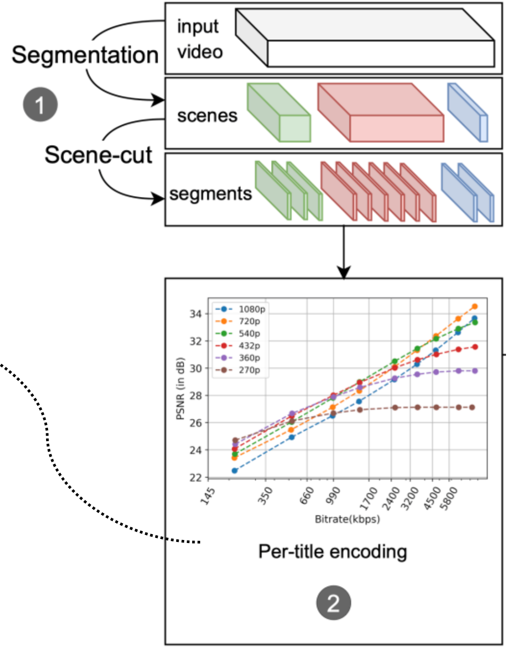
[**Amirpour**, et al, 2023, IEEE TSV]

Bitrate Ladder Construction

CPU vs GPU devices



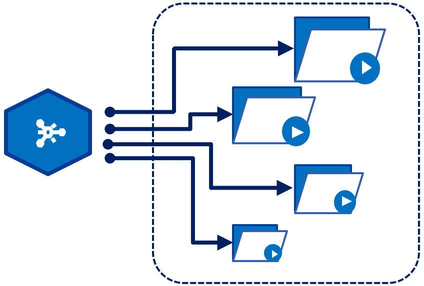
	Encoding Bitrate	Resolution
Rep. #01	145 kbps	640x360
Rep. #02	300 kbps	768x432
Rep. #03	600 kbps	960x540
Rep. #04	900 kbps	960x540
Rep. #05	1600 kbps	960x540
Rep. #06	2400 kbps	1280x720
Rep. #07	3400 kbps	1280x720
Rep. #08	4500 kbps	1920x1080
Rep. #09	5800 kbps	1920x1080
Rep. #10	8100 kbps	2560x1440
Rep. #11	11600 kbps	3840x2160
Rep. #12	16800 kbps	3840x2160



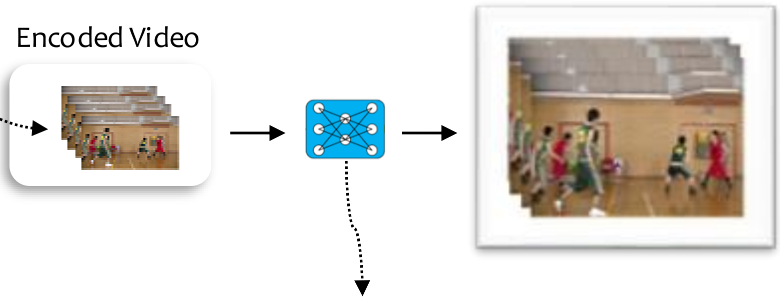
Backward Compatibility

Bitrate Ladder Construction

CPU vs GPU devices (DNN training)

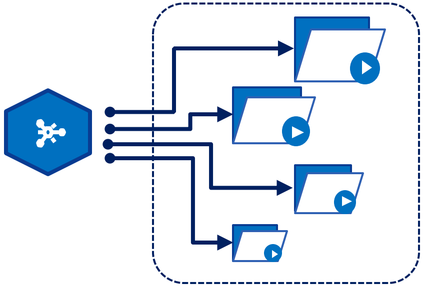


	Encoding Bitrate	Resolution
Rep.#01	145 kbps	640x360
Rep.#02	300 kbps	768x432
Rep.#03	600 kbps	960x540
Rep.#04	900 kbps	960x540
Rep.#05	1600 kbps	960x540
Rep.#06	2400 kbps	1280x720
Rep.#07	3400 kbps	1280x720
Rep.#08	4500 kbps	1920x1080
Rep.#09	5800 kbps	1920x1080
Rep.#10	8100 kbps	2560x1440
Rep.#11	11600 kbps	3840x2160
Rep.#12	16800 kbps	3840x2160

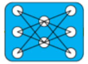


Bitrate Ladder Construction

CPU vs GPU devices

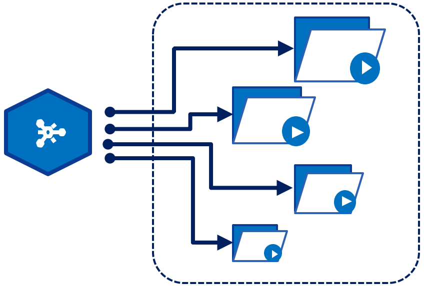


	Encoding Bitrate	Resolution
Rep. #01	145 kbps	640x360
Rep. #02	300 kbps	768x432
Rep. #03	600 kbps	960x540
Rep. #04	900 kbps	960x540
Rep. #05	1600 kbps	960x540
Rep. #06	2400 kbps	1280x720
Rep. #07	3400 kbps	1280x720
Rep. #08	4500 kbps	1920x1080
Rep. #09	5800 kbps	1920x1080
Rep. #10	8100 kbps	2560x1440
Rep. #11	11600 kbps	3840x2160
Rep. #12	16800 kbps	3840x2160



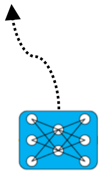
Bitrate Ladder Construction

CPU vs GPU devices (DNN Compression)



	Encoding Bitrate	Resolution
Rep.#01	145 kbps	640x360
Rep.#02	300 kbps	768x432
Rep.#03	600 kbps	960x540
Rep.#04	900 kbps	960x540
Rep.#05	1600 kbps	960x540
Rep.#06	2400 kbps	1280x720
Rep.#07	3400 kbps	1280x720
Rep.#08	4500 kbps	1920x1080
Rep.#09	5800 kbps	1920x1080
Rep.#10	8100 kbps	2560x1440
Rep.#11	11600 kbps	3840x2160
Rep.#12	16800 kbps	3840x2160

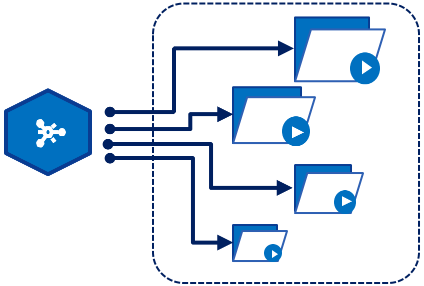
Compression



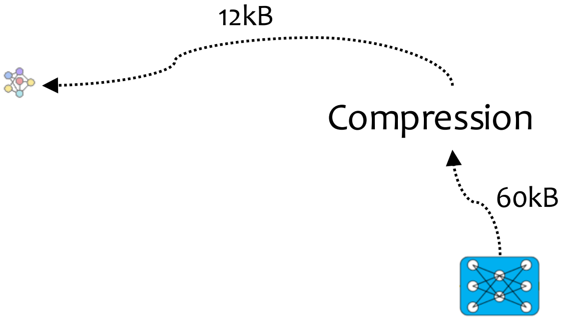
1- Pruning
2- Bit Precision
3- Quantization
4- ...

Bitrate Ladder Construction

CPU vs GPU devices (DNN Compression)

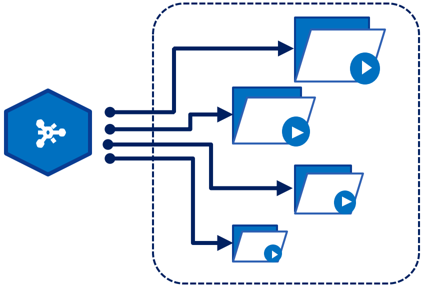


	Encoding Bitrate	Resolution
Rep.#01	145 kbps	640x360
Rep.#02	300 kbps	768x432
Rep.#03	600 kbps	960x540
Rep.#04	900 kbps	960x540
Rep.#05	1600 kbps	960x540
Rep.#06	2400 kbps	1280x720
Rep.#07	3400 kbps	1280x720
Rep.#08	4500 kbps	1920x1080
Rep.#09	5800 kbps	1920x1080
Rep.#10	8100 kbps	2560x1440
Rep.#11	11600 kbps	3840x2160
Rep.#12	16800 kbps	3840x2160



Bitrate Ladder Construction

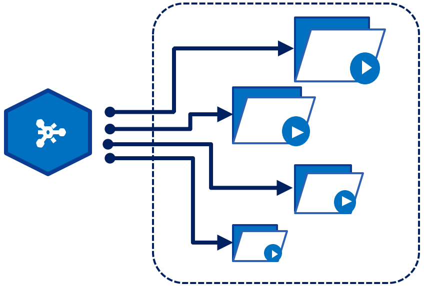
CPU vs GPU devices



	Encoding Bitrate	Resolution
Rep. #01	145 kbps	640x360
Rep. #02	300 kbps	768x432
Rep. #03	600 kbps	960x540
Rep. #04	900 kbps	960x540
Rep. #05	1600 kbps	960x540
Rep. #06	2400 kbps	1280x720
Rep. #07	3400 kbps	1280x720
Rep. #08	4500 kbps	1920x1080
Rep. #09	5800 kbps	1920x1080
Rep. #10	8100 kbps	2560x1440
Rep. #11	11600 kbps	3840x2160
Rep. #12	16800 kbps	3840x2160

Bitrate Ladder Construction

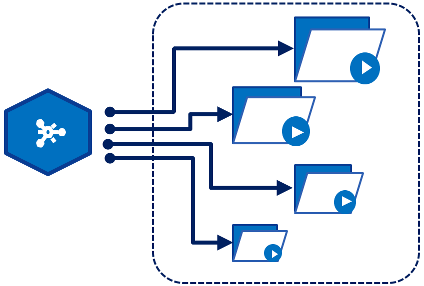
CPU vs GPU devices



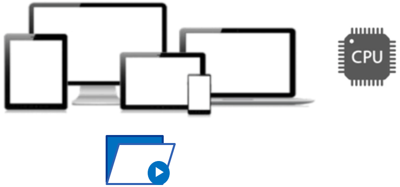
	Encoding Bitrate	Resolution
Rep.#01	145 kbps	640x360
Rep.#02	300 kbps	768x432
Rep.#03	600 kbps	960x540
Rep.#04	900 kbps	960x540
Rep.#05	1600 kbps	960x540
Rep.#06	2400 kbps	1280x720
Rep.#07	3400 kbps	1280x720
Rep.#08	4500 kbps	1920x1080
Rep.#09	5800 kbps	1920x1080
Rep.#10	8100 kbps	2560x1440
Rep.#11	11600 kbps	3840x2160
Rep.#12	16800 kbps	3840x2160

Bitrate Ladder Construction

CPU vs GPU devices



	Encoding Bitrate	Resolution
Rep. #01	145 kbps	640x360
Rep. #02	300 kbps	768x432
Rep. #03	600 kbps	960x540
Rep. #04	900 kbps	960x540
Rep. #05	1600 kbps	960x540
Rep. #06	2400 kbps	1280x720
Rep. #07	3400 kbps	1280x720
Rep. #08	4500 kbps	1920x1080
Rep. #09	5800 kbps	1920x1080
Rep. #10	8100 kbps	2560x1440
Rep. #11	11600 kbps	3840x2160
Rep. #12	16800 kbps	3840x2160

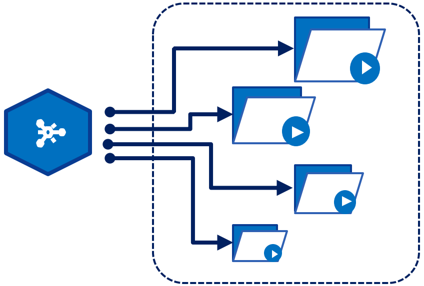


UNIVERSITÄT KLAGENFURT

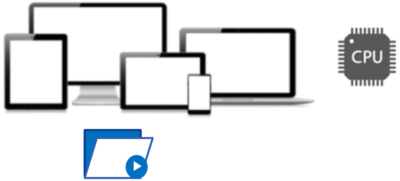
ATHENA
Christian Doppler (CD) Laboratory

Bitrate Ladder Construction

CPU vs GPU devices

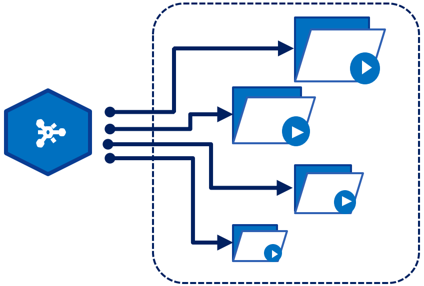


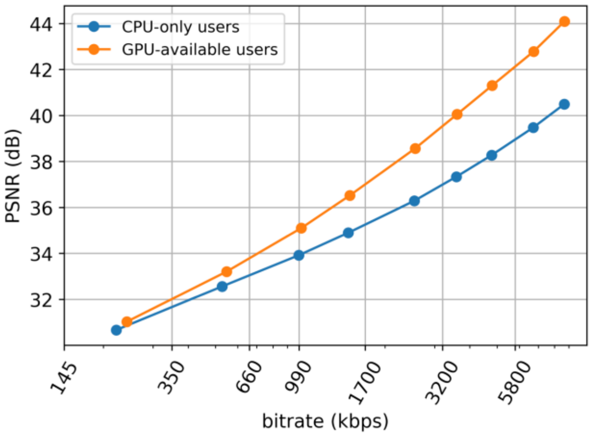
	Encoding Bitrate	Resolution
Rep. #01	145 kbps	640x360
Rep. #02	300 kbps	768x432
Rep. #03	600 kbps	960x540
Rep. #04	900 kbps	960x540
Rep. #05	1600 kbps	960x540
Rep. #06	2400 kbps	1280x720
Rep. #07	3400 kbps	1280x720
Rep. #08	4500 kbps	1920x1080
Rep. #09	5800 kbps	1920x1080
Rep. #10	8100 kbps	2560x1440
Rep. #11	11600 kbps	3840x2160
Rep. #12	16800 kbps	3840x2160



Bitrate Ladder Construction

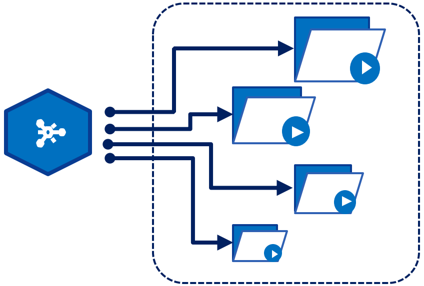
CPU vs GPU devices (Compression Efficiency)

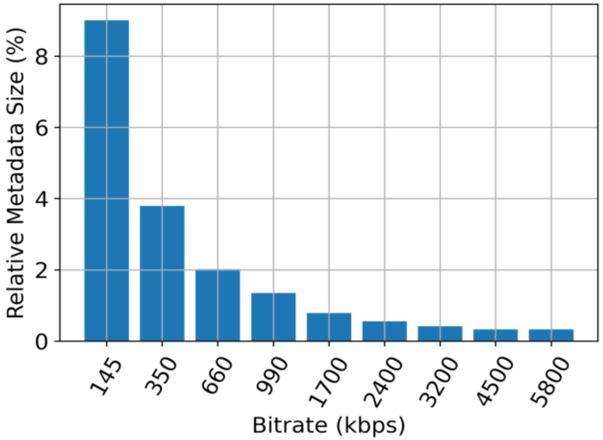




Bitrate Ladder Construction

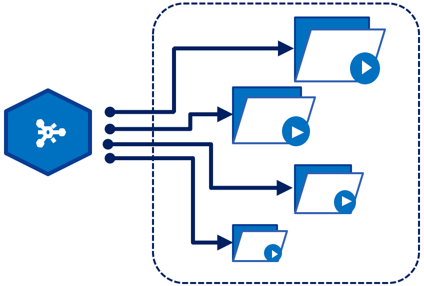
CPU vs GPU devices





Bitrate Ladder Construction

CPU vs GPU devices



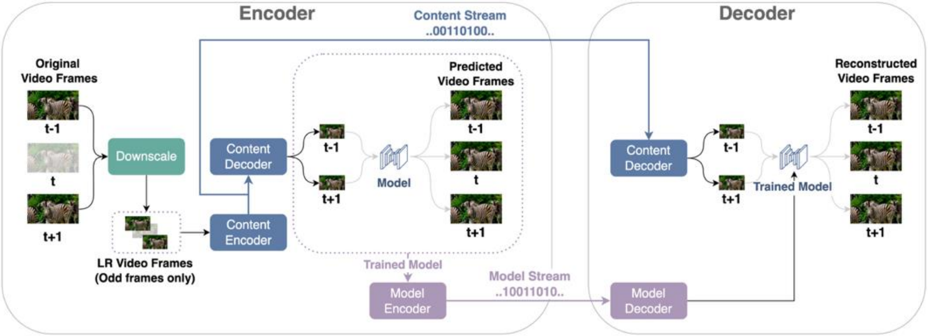


Fig. 1: Overview of proposed video compression pipeline. Videos are encoded into two bitstreams: content stream and model stream. Content stream encodes LSTR video with existing codec. The model stream encodes a small portion of parameter updates for the DNN model which is capable of interpolating and super-resolve decompressed video. Note that we only show two input LR frames from a long sequence in this figure for better illustration.

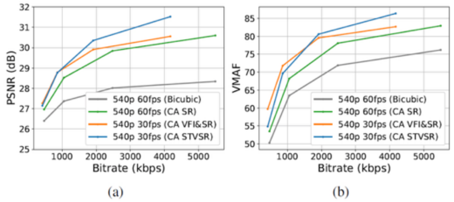


Fig. 3: RD curves using (a) PSNR and (b) VMAF of content-aware (CA) SR, CA VFI&SR, and CA STVSR methods for video #12.

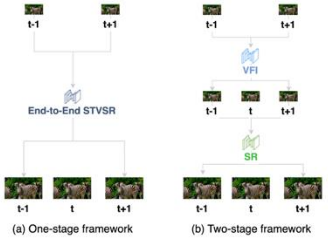
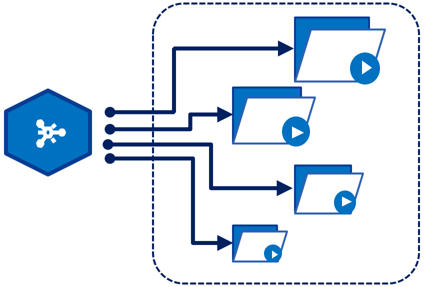


Fig. 2: Illustration of STVSR frameworks.

Bitrate Ladder Construction

CPU vs GPU devices



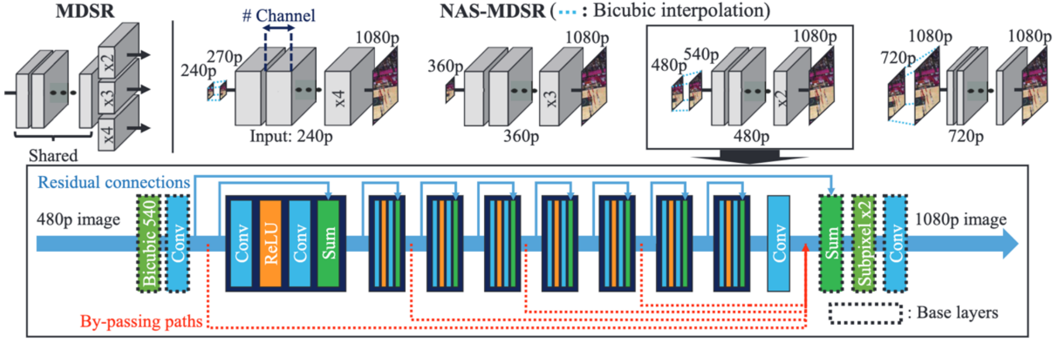
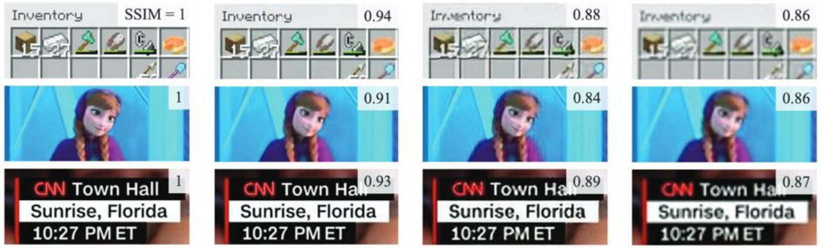


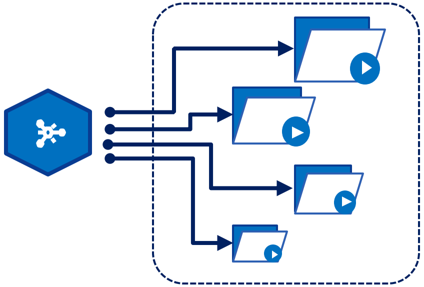
Figure 4: Super-resolution DNN architectures: MDSR vs. NAS-MDSR



(a) Original (1080p) (b) Content-aware DNN (c) Content-agnostic DNN (d) 240p

Bitrate Ladder Construction

CPU vs GPU devices



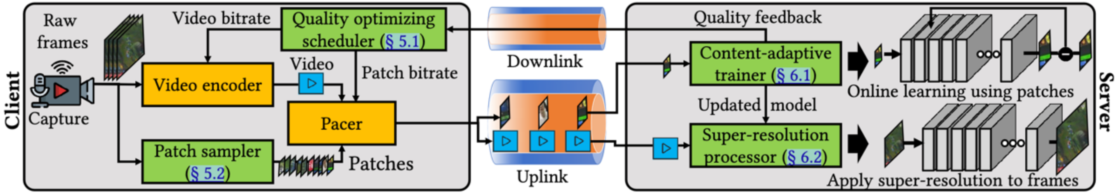


Figure 3: LiveNAS System Overview

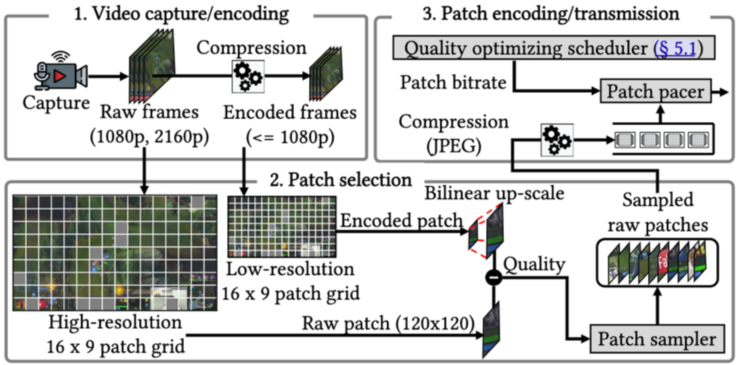
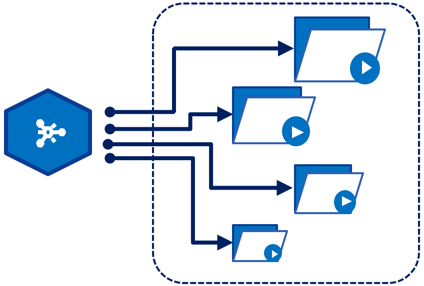


Figure 7: Patch Selection Overview

Bitrate Ladder Construction

CPU vs GPU devices



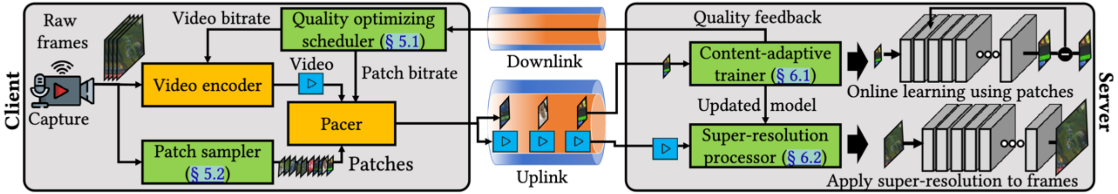


Figure 3: LiveNAS System Overview

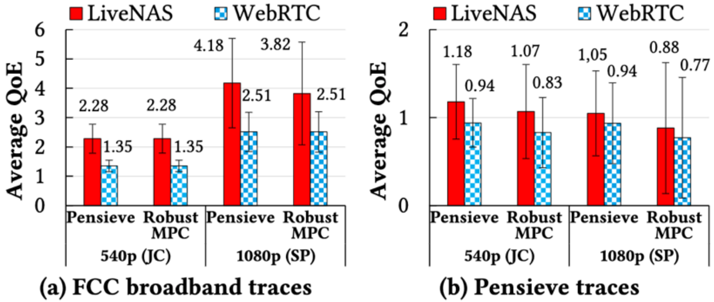
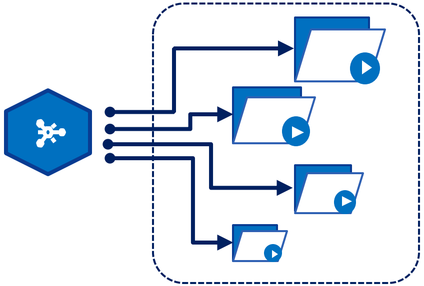


Figure 20: Impact on viewer QoE at distribution-side.

Bitrate Ladder Construction

CPU vs GPU devices



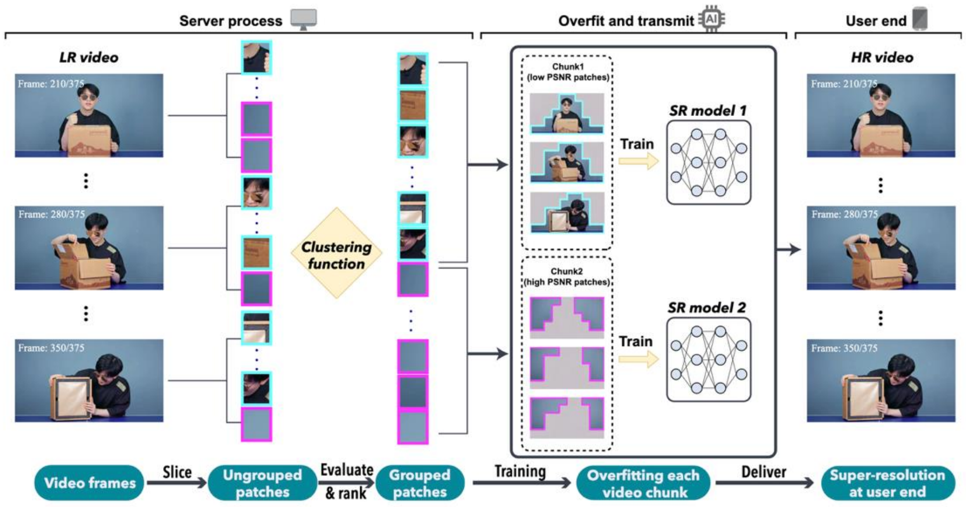
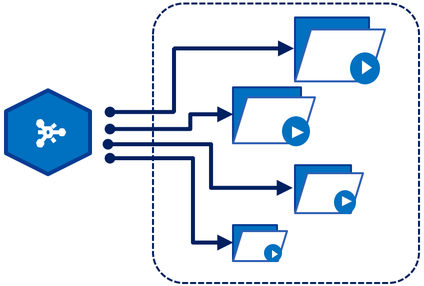


Figure 1. Patch PSNR heatmap of two frames in a 15s video when super-resolved by a general WDSR model. A clear boundary shows that PSNR is strongly related to video content.

Bitrate Ladder Construction

CPU vs GPU devices



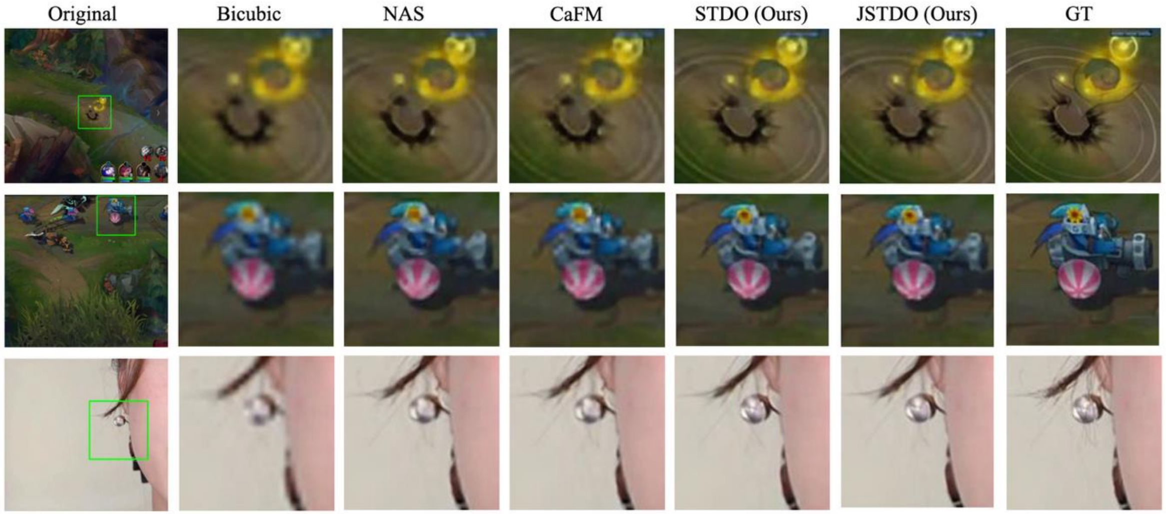
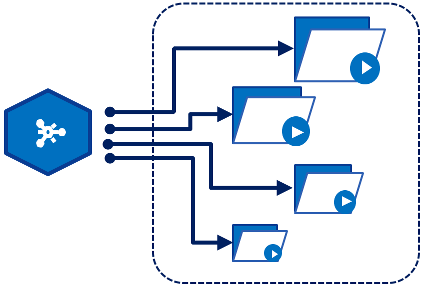
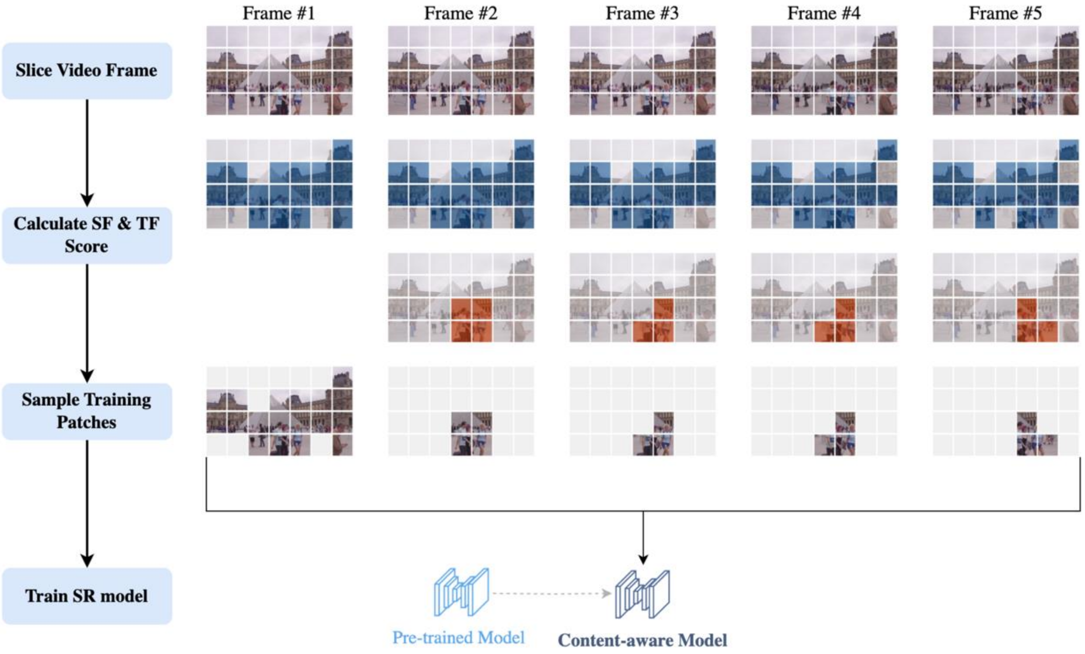


Figure 4. Super-resolution quality comparison with random video frame using STDO and JSTDO with baseline methods.

Bitrate Ladder Construction

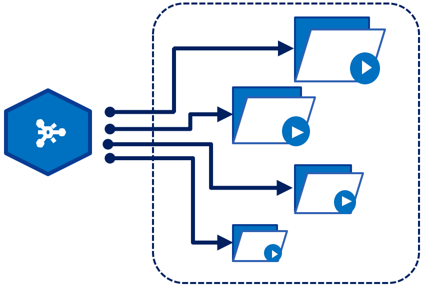
CPU vs GPU devices

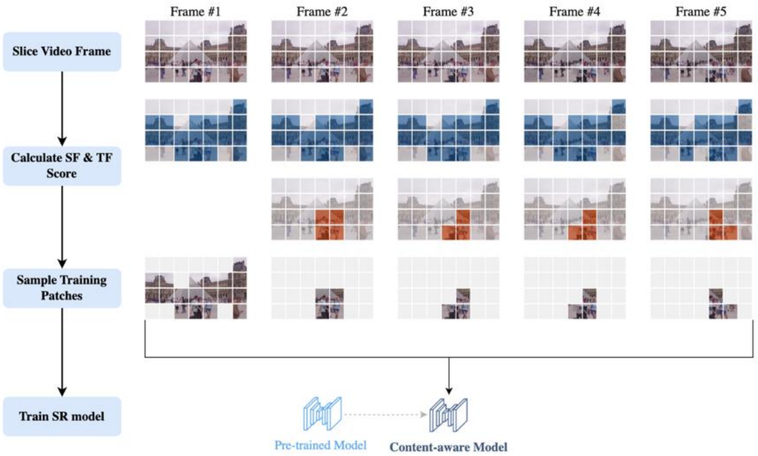


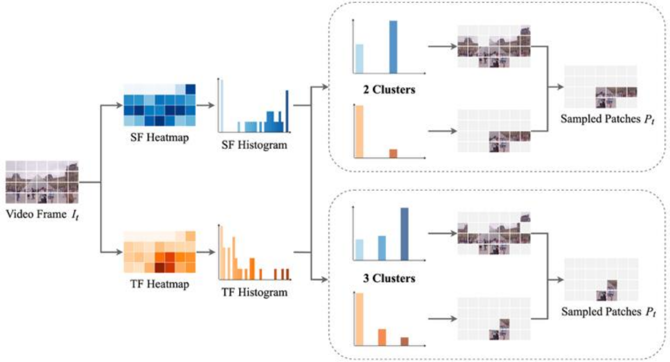


Bitrate Ladder Construction

CPU vs GPU devices

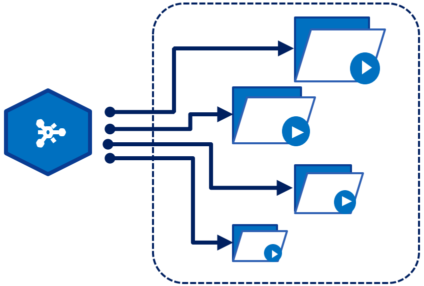




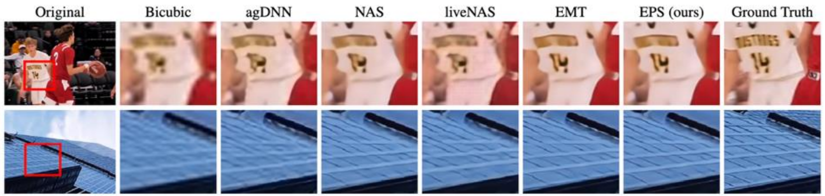


Bitrate Ladder Construction

CPU vs GPU devices

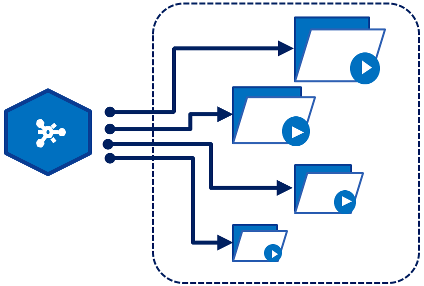


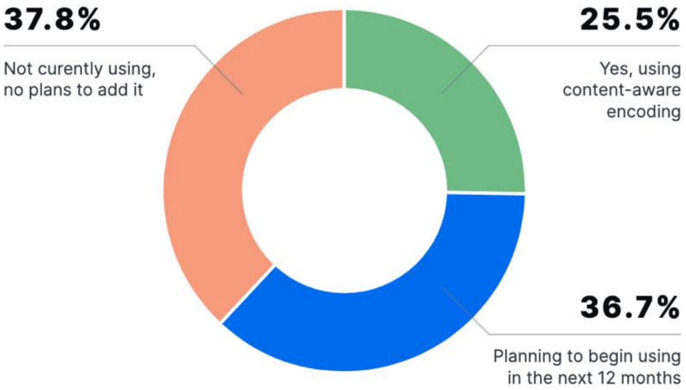
Scale Factor	Method	Patches	FSRCNN PSNR	FSRCNN VMAF	ESPCN PSNR	ESPCN VMAF	CARN PSNR	CARN VMAF	WDSR PSNR	WDSR VMAF
× 2	agDNN [30]	0 %	26.15	79.72	32.40	82.12	35.41	88.90	35.83	89.49
	NAS [31]	100.00 %	34.74	84.61	33.96	83.46	36.12	89.14	36.72	90.10
	LiveNAS [13]	17.50 %	33.60	80.01	33.27	83.14	35.76	88.95	36.27	89.47
	EMT [17]	17.50 %	34.26	85.13	34.49	84.52	35.89	89.03	36.71	90.02
	EPS (ours)	17.50 %	**34.85**	**85.71**	**34.61**	**85.08**	**36.42**	**89.91**	**36.78**	**90.19**
× 4	agDNN [30]	0 %	24.15	47.95	29.44	43.01	30.86	64.09	30.69	63.92
	NAS [31]	100.00 %	29.99	51.75	29.90	50.77	**31.89**	**69.71**	**31.87**	**69.66**
	LiveNAS [13]	26.78 %	29.27	48.37	29.65	47.75	30.93	64.67	31.20	67.53
	EMT [17]	26.78 %	30.09	52.71	29.86	51.35	31.21	68.34	31.48	69.07
	EPS (ours)	26.78 %	**30.16**	**54.17**	**30.06**	**52.14**	31.77	69.25	31.79	69.55



Bitrate Ladder Construction

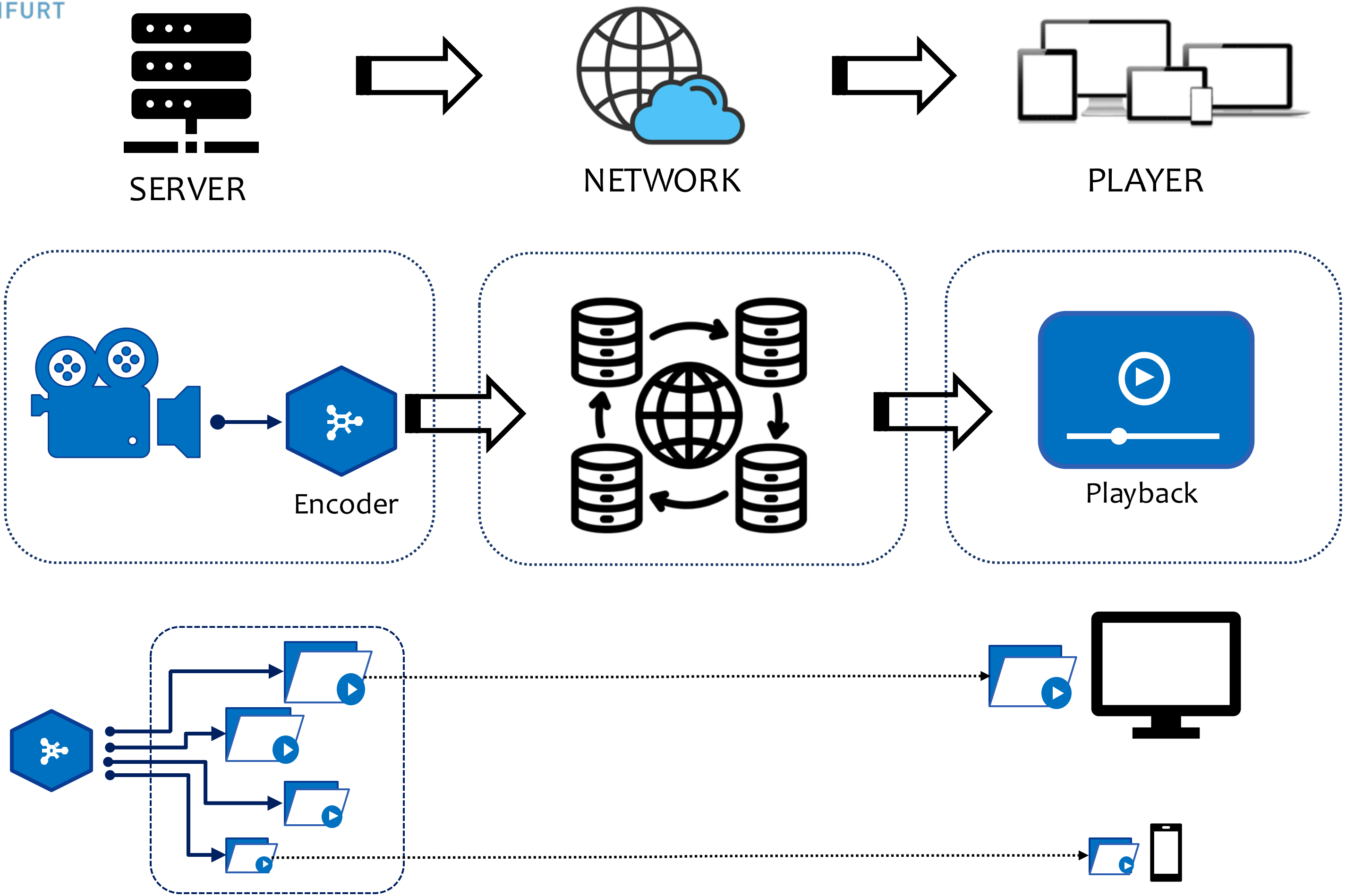
CPU vs GPU devices





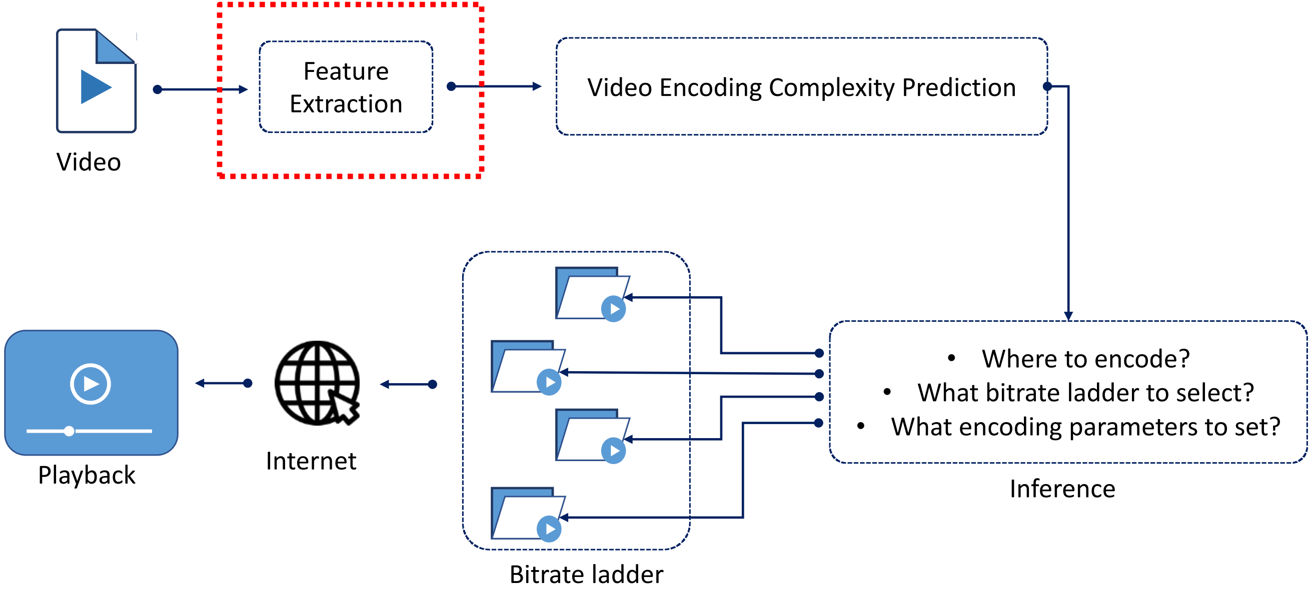
Are you using or planning to use content-aware encoding technology (i.e., Per-Title)?

The majority of video professionals aren't using content-aware encoding but adoption plans continue to grow. Even so, we've seen similar numbers for the past couple of years, indicating that respondents aren't following through on their plans to implement it. We attribute this to economic pressures, despite the fact that content-aware encoding would deliver cost efficiencies in the long run.


SERVER
NETWORK
PLAYER
Encoder
Playback

Bitrate Ladder Construction

Feature Extraction for Live Video Streaming



Bitrate Ladder Construction

SI-TI

- The state-of-the-art spatial and temporal complexity feature is **SI-TI.**

Sobel Filter

Orginal Frame | Spatial Compelxity | Temporal Complexity

$$SI = stdspace(Sobel(F_t))$$

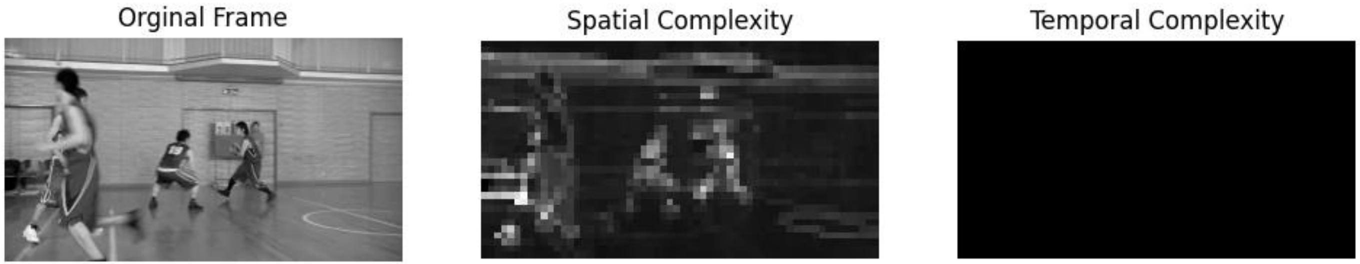
$$TI = stdspace[F_t(i, j) - F_{t-1}(i, j)]$$

- The correlation with ground truth spatial and temporal complexity is low!!!

Bitrate Ladder Construction

Video Complexity Analzyer (VCA)

- VCA introduces DCT-based spatial-temporal complexity features



$$E_{DCT}(c) = \sum_{i=1}^{w}\sum_{j=1}^{h} e^{\left(\frac{i\cdot j}{w\cdot h}\right)^2-1} \left|DCT_c(i-1, j-1, c)\right|$$

$$h = \frac{1}{C}\sum_{c=2}^{C}\frac{1}{w^2} SAD(E_{DCT}(t, c), E_{DCT}(t-1, c))$$

- Spatial complexity correlation has increased, but the temporal complexity remains low.

[Menon, **Amirpour**, et al, 2022, ACM MMSys]

Bitrate Ladder Construction

Video Complexity Analyzer (VCA)

➢ The state-of-the-art spatial and temporal complexity feature is **SI-TI** [1,2].

[1] ITU-T P910 Subjective video quality assessment methods for multimedia applications.
[2] SITI source code: https://github.com/Telecommunication-Telemedia-Assessment/SITI

Bitrate Ladder Construction

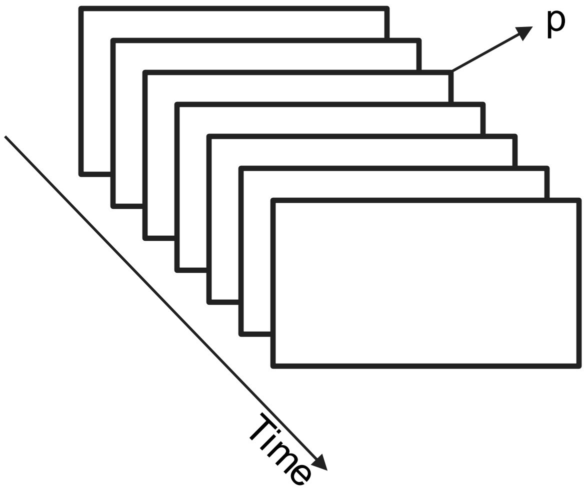
Video Complexity Analyzer (VCA)

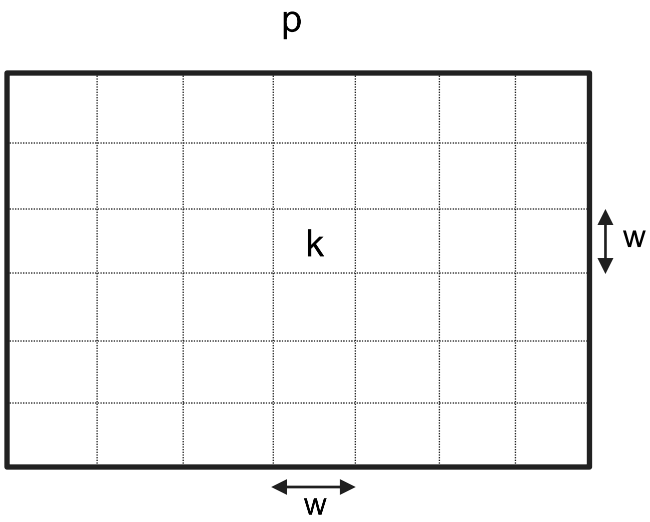
- The state-of-the-art spatial and temporal complexity feature is **SI-TI** [1,2].
- The correlation with encoding bitrate and encoding time is low!

[1] ITU-T P910 Subjective video quality assessment methods for multimedia applications.
[2] SITI source code: https://github.com/Telecommunication-Telemedia-Assessment/SITI

Bitrate Ladder Construction

Video Complexity Analyzer (VCA)





$$H_{p,k} = \sum_{i=0}^{w-1} \sum_{j=0}^{w-1} e^{\left|\left(\frac{ij}{w^2}\right)^2 - 1\right|} |DCT(i,j)|$$

$DCT(i,j)$ is the $(i,j)^{th}$ DCT component, $i + j > 0$

0, otherwise

Bitrate Ladder Construction

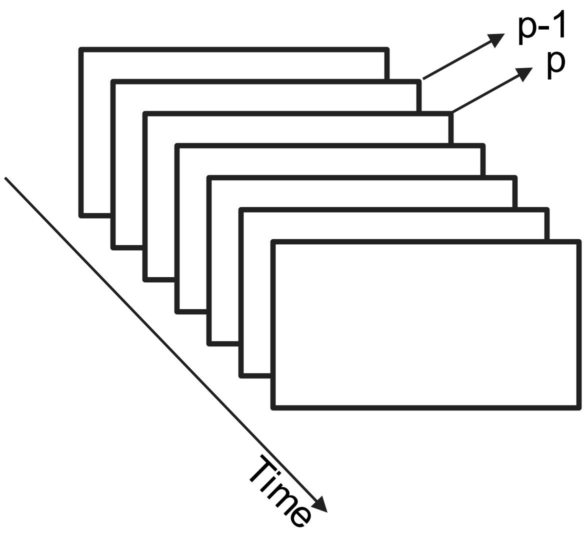
Video Complexity Analyzer (VCA)

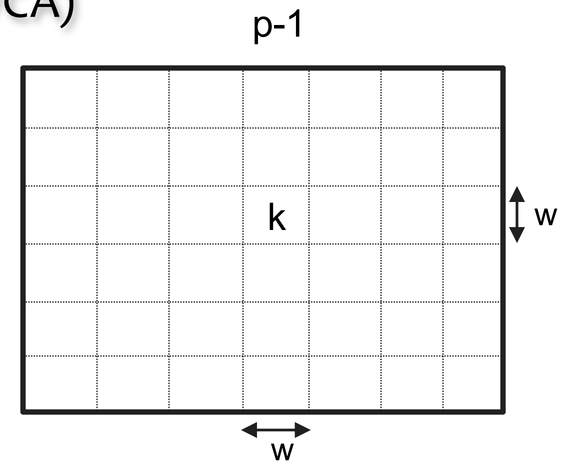
$$E = \sum_{k=0}^{C-1} \frac{H_{p,k}}{C.w^2}$$

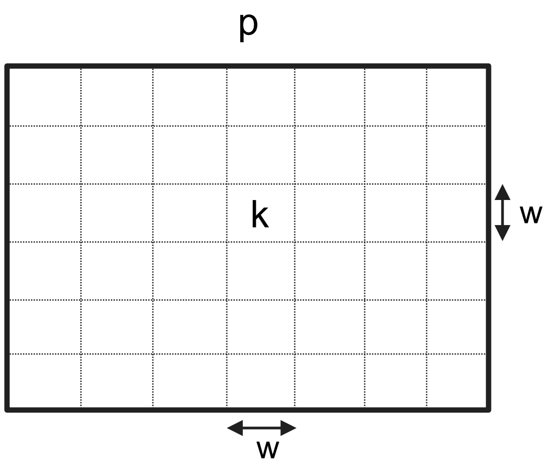
Where C represents the number of blocks per frame.

Bitrate Ladder Construction

Video Complexity Analyzer (VCA)



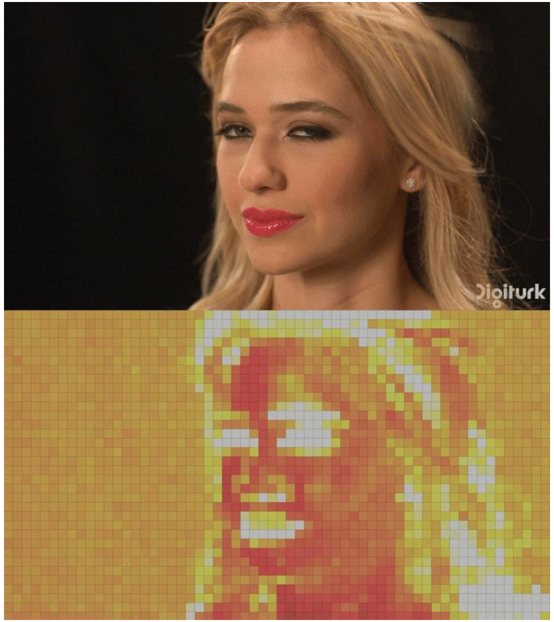




$$h = \sum_{k=0}^{C-1} \frac{SAD(H_{p,k}, H_{p-1,k})}{C.w^2}$$

Bitrate Ladder Construction

Video Complexity Analyzer (VCA)

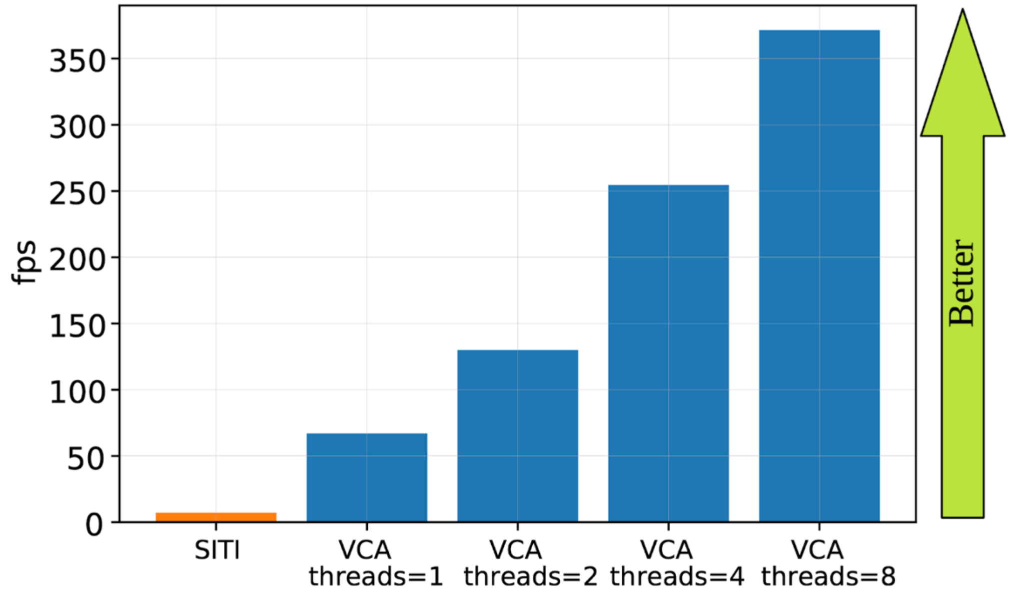


Spatial Complexity (E)

Temporal Complexity (h)

Bitrate Ladder Construction

Video Complexity Analyzer (VCA)

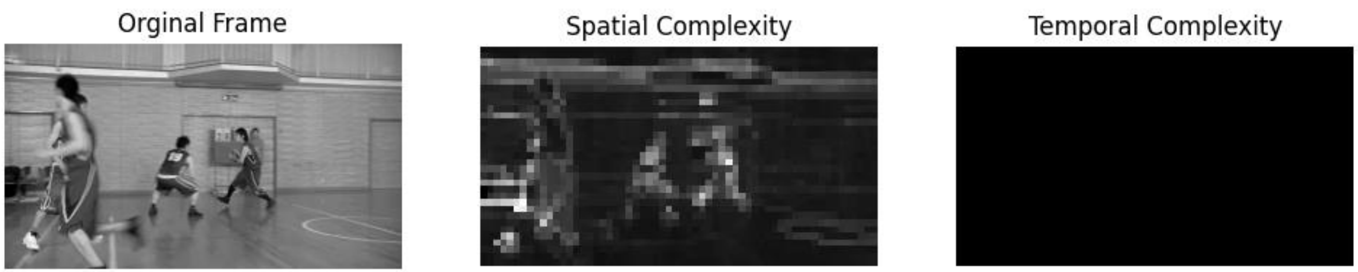


Average time to compute E-h for 2160p (with x86 SIMD and multi-threading).

Bitrate Ladder Construction

Enhanced Video Complexity Analyzer (EVCA)

- EVCA refines the definition of the temporal complexity feature



$$E_{DCT}(c) = \sum_{i=1}^{w}\sum_{j=1}^{h} e^{\left(\frac{i \cdot j}{w \cdot h}\right)^2 - 1} \left|DCT_c(i-1, j-1, c)\right|$$

$$TC_c = \sum_{i=1}^{w}\sum_{j=1}^{h} e^{\left(\frac{i \cdot j}{w \cdot h}\right)^2 - 1} \left|DCT(i-1, j-1, c) - DCT(i-1, j-1, c-1)\right|$$

- Temporal complexity correlation has increased.

[**Amirpour**, et al, 2024, ACM MMSys]

UNIVERSITÄT KLAGENFURT

ATHENA Christian Doppler (CD) Laboratory

Bitrate Ladder Construction

Enhanced Video Complexity Analyzer (EVCA)

Feature Extraction		
SITI	VCA	EVCA
1538 fps	1502 fps	1250 fps

Table 2: Performance of spatial complexity features.

Metric	Features		
	SI (SITI)	E (VCA)	SC (EVCA)
PCC	72	**93**	**93**
SRCC	78	**95**	**95**

Table 3: Performance of temporal complexity features.

Metric	Features		
	TI (SITI)	h (VCA)	TC (EVCA)
PCC	62	61	**77**
SRCC	73	77	**85**

[**Amirpour**, et al, 2024, ACM MMSys]

Bitrate Ladder Construction

Feature Extraction for Live Video Streaming

SITI VCA EVCA

Bitrate Ladder Construction

Feature Extraction for Live Video Streaming

Unsupervised

SITI VCA EVCA

Bitrate Ladder Construction

Feature Extraction for Live Video Streaming

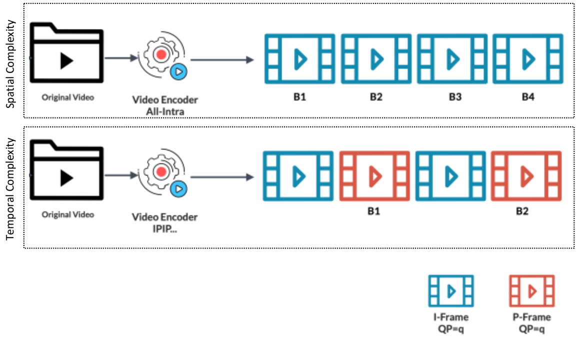
Unsupervised

SITI VCA EVCA

Supervised

Labels?

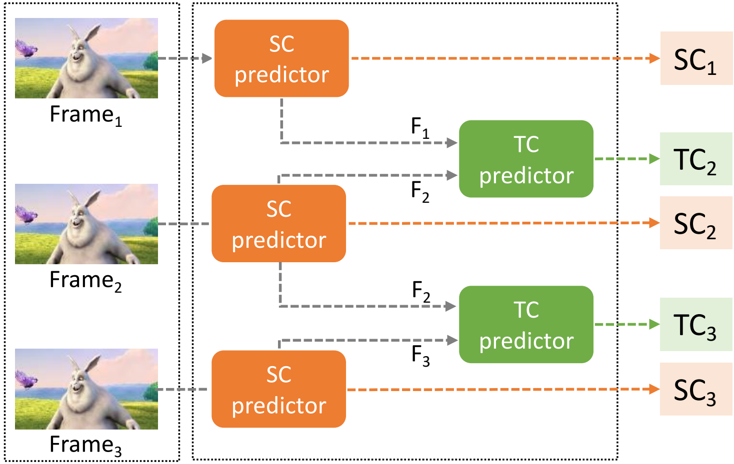
DeepVCA



[**Amirpour**, et al, 2024, IEEE TCSVT]

Bitrate Ladder Construction

Deep Video Complexity Analyzer (DeepVCA)



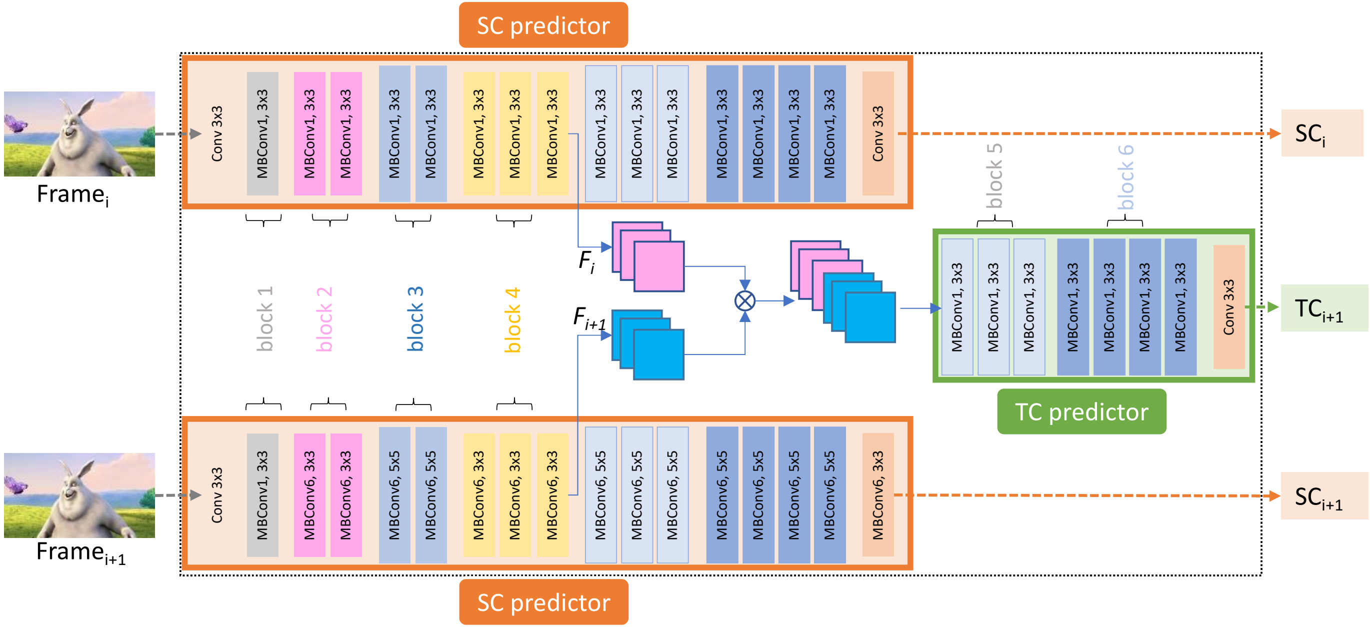
[**Amirpour**, et al, 2024, IEEE TCSVT]

UNIVERSITÄT KLAGENFURT

ATHENA
Christian Doppler (CD) Laboratory

Bitrate Ladder Construction

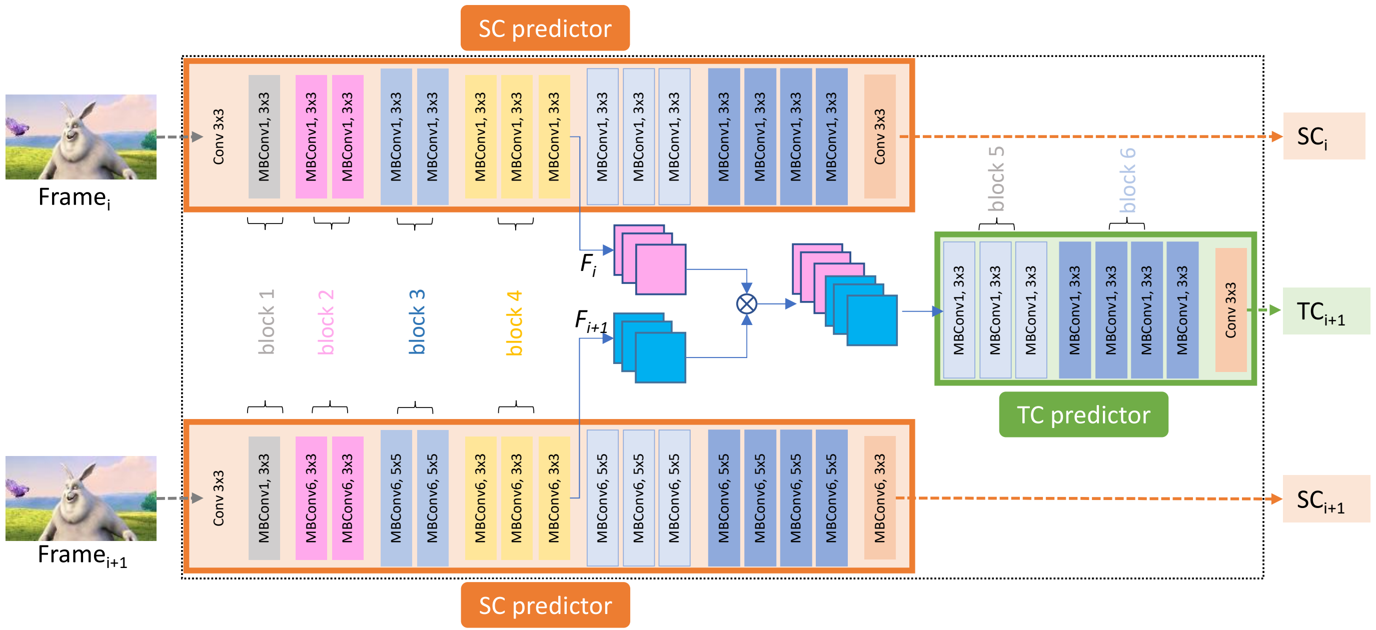
Deep Video Complexity Analyzer (DeepVCA)



[**Amirpour**, et al, 2024, IEEE TCSVT]

Bitrate Ladder Construction

Deep Video Complexity Analyzer (DeepVCA)



[**Amirpour**, et al, 2024, IEEE TCSVT]

Bitrate Ladder Construction

Deep Video Complexity Analyzer (DeepVCA)

TABLE I: Comparison of spatial complexity metrics.

	Metric	x265, QP=22 PCC	x265, QP=22 SRCC	x265, QP=22 KRCC	x265, QP=37 PCC	x265, QP=37 SRCC	x265, QP=37 KRCC	x265, QP=12 PCC	x265, QP=12 SRCC	x265, QP=12 KRCC	x264, QP=22 PCC	x264, QP=22 SRCC	x264, QP=22 KRCC
Unsupervised	SI	0.72	0.76	0.56	0.82	0.87	0.69	0.71	0.76	0.56	0.74	0.80	0.61
	E	0.91	0.93	0.78	**0.95**	0.95	0.83	0.91	0.93	0.78	0.93	0.95	0.82
Supervised (*DeepVCA*)	AlexNet	0.83	0.83	0.63	0.81	0.83	0.64	0.83	0.83	0.65	0.83	0.83	0.64
	VGG11	0.92	0.92	0.76	0.89	0.91	0.74	0.92	0.92	0.76	0.92	0.92	0.76
	ResNet-18	0.94	0.95	0.81	0.94	0.96	0.83	0.94	0.95	0.81	0.95	0.95	0.82
	MobileNetV2	**0.97**	**0.98**	0.87	**0.95**	**0.96**	0.83	**0.97**	0.97	0.86	0.95	0.95	0.82
	EfficientNet-b0	**0.97**	**0.98**	**0.88**	**0.95**	**0.96**	**0.84**	**0.97**	**0.98**	**0.87**	**0.97**	**0.98**	**0.88**

TABLE II: Comparison of temporal complexity metrics.

	Metric	x265, QP=22 PCC	x265, QP=22 SRCC	x265, QP=22 KRCC	x265, QP=37 PCC	x265, QP=37 SRCC	x265, QP=37 KRCC	x265, QP=12 PCC	x265, QP=12 SRCC	x265, QP=12 KRCC	x264, QP=22 PCC	x264, QP=22 SRCC	x264, QP=22 KRCC
Unsupervised	TI	0.43	0.46	0.32	0.21	0.39	0.27	0.57	0.53	0.68	0.42	0.45	0.32
	h	0.46	0.52	0.37	0.24	0.47	0.32	0.58	0.56	0.41	0.45	0.52	0.37
Supervised (*DeepVCA*)	MobileNetV2 (block 11)	0.79	0.83	0.64	0.49	0.64	0.46	**0.84**	0.86	0.68	0.77	0.81	0.63
	MobileNetV2 (block 4)	0.80	0.83	0.65	0.54	0.65	0.47	**0.84**	0.87	0.69	0.79	082	0.64
	EfficientNet-b0 (block 6)	0.77	0.81	0.63	0.49	0.59	0.43	0.78	0.84	0.66	0.74	0.79	0.61
	EfficientNet-b0 (block 4)	0.81	0.84	0.66	0.55	0.65	0.48	0.83	0.86	0.68	0.79	0.83	0.65
	EfficientNet-b0 (block 2)	0.81	0.85	0.68	0.57	0.67	0.50	0.83	**0.87**	0.68	0.80	0.84	0.67
	3D CNN: R(2+1)D-18	**0.84**	**0.86**	**0.69**	**0.64**	**0.73**	**0.54**	**0.84**	**0.87**	**0.69**	**0.83**	**0.85**	**0.68**

[**Amirpour**, et al, 2024, IEEE TCSVT]

Bitrate Ladder Construction

Open-source Software

	Device	Description
VCA	CPU	Optimized software
EVCA	GPU	Enhanced temporal complexity
DeepVCA	GPU	Enhanced Temporal complexity

Bitrate Ladder Construction

Motion-Inspired Image Complexity Metric

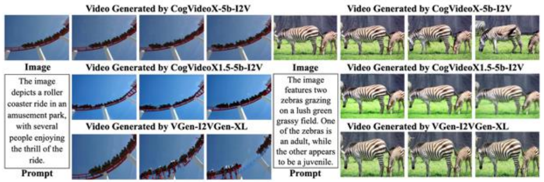
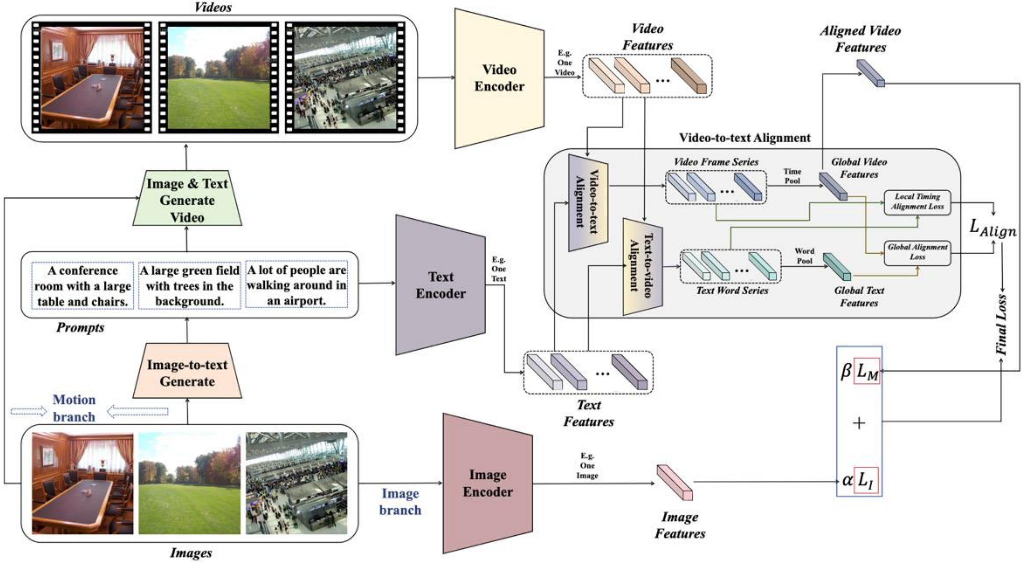
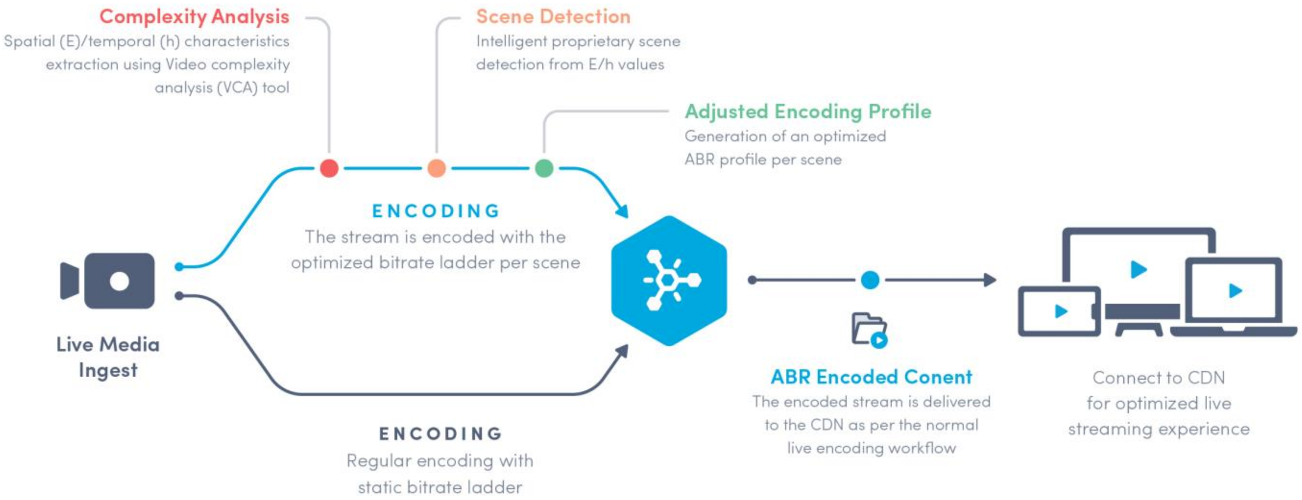


Figure 5: The examples of images in IC9600 database, and corresponding text prompts generated by InstructBLIP [11] (Vicuna-7B). The video frames are sampled from three state-of-the-art image-to-video models (i.e., CogVideoX [66], CogVideoX1.5 [66], and Vgen [69].)



Bitrate Ladder Construction

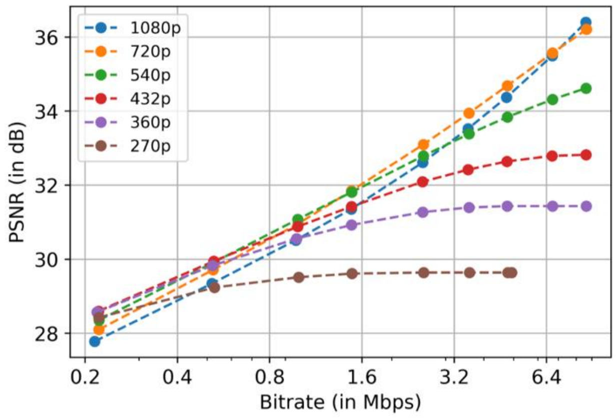
Live Video Streaming

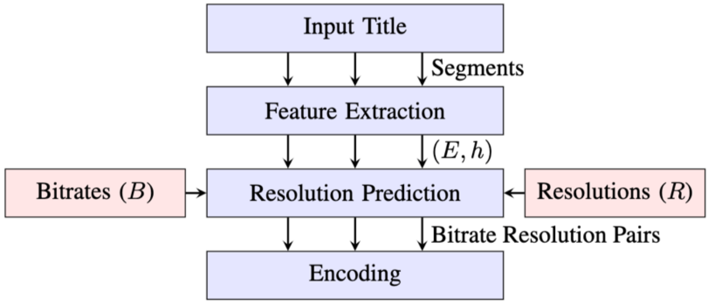


[**Amirpour**, et al, 2022, IEEE ICME]

Bitrate Ladder Construction

Live Video Streaming





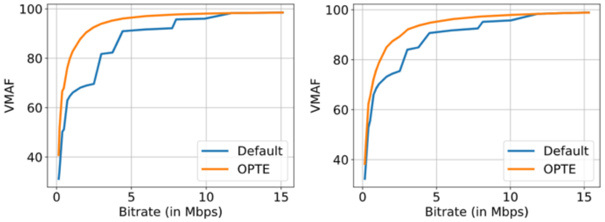
[**Amirpour**, et al, 2022, IEEE ICME]

UNIVERSITÄT KLAGENFURT

ATHENA Christian Doppler (CD) Laboratory

Bitrate Ladder Construction

Live Video Streaming



(a) *BundNightScape* (b) *Bunny*

Table 1: Results of OPTE against fixed bitrate ladder approach.

Dataset	Video	f	SI	TI	E	h	$\|\| s_G - s \|\|_2$	BDR_V	BDR_P
MCML	Bunny	30	23.38	6.43	23.03	4.88	0.01	-39.48%	-32.25%
MCML	Characters	30	50.43	29.85	41.44	29.21	0.04	-51.90%	-68.81%
MCML	Crowd	30	33.76	10.13	33.11	12.22	0.01	-29.82%	-14.18%
MCML	Dolls	30	16.88	19.91	10.47	0.27	0.02	-1.43%	-8.49%
SJTU	BundNightScape	30	48.82	7.06	54.90	11.62	0.02	-61.22%	-60.86%
SJTU	Fountains	30	43.37	11.42	60.90	23.02	0.02	-32.93%	-8.49%
SJTU	TrafficFlow	30	33.57	13.80	58.93	15.83	0.02	-50.54%	-40.90%
SJTU	TreeShade	30	52.88	5.29	80.19	8.83	0.01	-47.76%	-38.55%
VQEG	CrowdRun	50	50.77	22.33	96.55	33.33	0.01	-8.50%	-1.90%
VQEG	DucksTakeOff	50	47.77	15.10	119.12	30.88	0.01	-2.99%	-2.79%
VQEG	IntoTree	50	24.41	12.09	74.45	21.95	0.03	-26.50%	-5.75%
VQEG	OldTownCross	50	29.66	11.62	92.75	22.06	0.02	-30.91%	-22.53%
VQEG	ParkJoy	50	62.78	27.00	102.80	52.15	0.02	-12.08%	-2.62%
JVET	CatRobot	60	44.45	11.84	56.36	14.25	0.01	-13.43%	-5.95%
JVET	DaylightRoad2	60	40.51	16.21	66.40	20.13	0.02	-27.52%	-9.35%
JVET	FoodMarket4	60	38.26	17.68	50.71	20.71	0.02	-18.11%	-3.74%
Average							0.02	**-28.45%**	**-20.45%**

[**Amirpour**, et al, 2022, IEEE ICME]

Bitrate Ladder Construction

Live Video Streaming

1.6 Mbps
1080p

1.6 Mbps
540p

[**Amirpour**, et al, 2022, IEEE ICME]

Bitrate Ladder Construction

Live Video Streaming

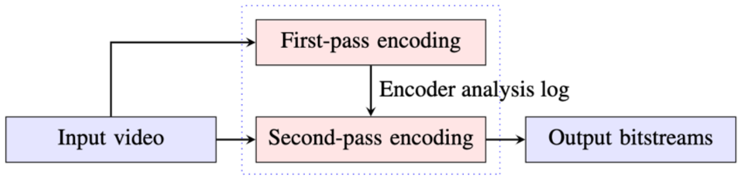


Fig. 1: Two-pass encoding architecture.

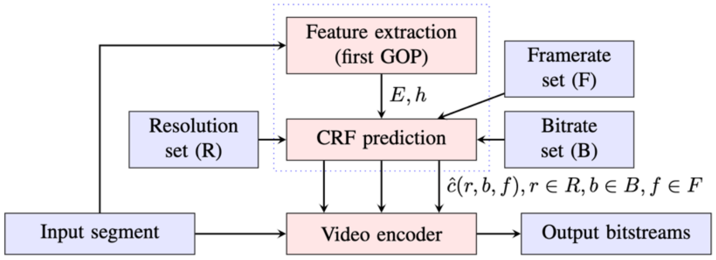


Fig. 2: `ETPS` architecture. The dashed blue line represents the first-pass while the video encoder block represents the second-pass encode.

Bitrate Ladder Construction

Live Video Streaming

Table 1: Results of `ETPS` against Constant Bitrate (CBR) encoding and two-pass average bitrate encoding schemes.

Dataset	Video	E	h	$\|\| c_G - \hat{c} \|\|_2$	τ_p (in ms)	Constant bitrate (CBR)			Two-pass average bitrate		
						BDR_P	BDR_V	ΔT	BDR_P	BDR_V	ΔT
MCML [18]	Basketball	15.75	9.69	0.86	8	-17.80%	-14.36%	0.43%	-0.48%	0.50%	-45.15%
MCML	Bunny	23.03	4.89	1.58	7	-9.92%	-9.09%	0.78%	0.04%	-1.14%	-43.88%
MCML	Crowd	33.11	7.03	1.58	7	-7.63%	-1.47%	0.56%	0.75%	1.05%	-43.34%
MCML	Dolls	19.91	12.22	0.86	8	-5.73%	-4.05%	1.01%	1.21%	1.97%	-45.26%
MCML	Flowers	12.01	10.47	0.50	7	-8.44%	-7.12%	-0.59%	0.38%	0.49%	-43.90%
MCML	Park	26.07	27.34	0.71	9	-5.16%	-2.62%	1.10%	0.16%	0.87%	-44.27%
SJTU [19]	BundNightScape	54.90	11.62	1.93	8	-14.63%	-13.51%	-1.62%	-1.11%	-1.27%	-40.91%
SJTU	CampfireParty	51.50	42.38	1.58	8	-4.66%	-3.66%	-1.72%	-0.51%	-0.35%	-43.04%
SJTU	Runners	104.30	17.44	1.32	7	-8.14%	-2.44%	-1.34%	-0.77%	-1.08%	-42.69%
SJTU	TallBuildings	94.35	7.70	1.65	8	-17.30%	-12.54%	-1.02%	0.89%	0.59%	-43.37%
SJTU	TrafficAndBuilding	60.54	11.55	1.87	8	-15.41%	-9.91%	-0.43%	0.06%	-2.42%	-43.96%
SJTU	TrafficFlow	58.93	8.61	1.32	8	-15.66%	-1.17%	0.74%	1.19%	1.43%	-43.86%
SJTU	TreeShade	80.19	15.83	0.50	8	-11.82%	-6.85%	-0.79%	0.52%	0.59%	-44.75%
SJTU	Wood	114.26	8.83	0.71	8	-10.14%	-21.58%	0.52%	1.34%	1.88%	-44.50%
Average				1.02	8	**-10.89%**	**-8.60%**	**0.65%**	**0.26%**	**0.38%**	**-43.78%**

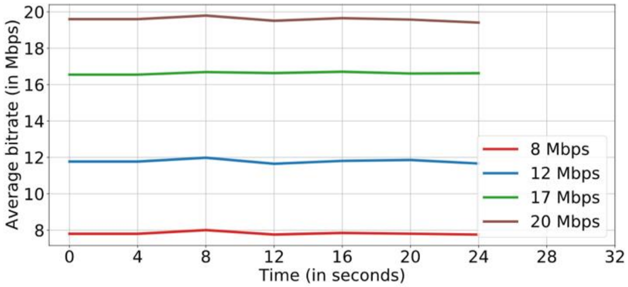


Fig. 4: The average bitrate in a streaming session where MCML sequences (*cf.* Table 1) are encoded using `ETPS` and are streamed at intervals of four seconds for target average bitrates 8, 12, 17, 20 Mbps.

Bitrate Ladder Construction

Live Video Streaming

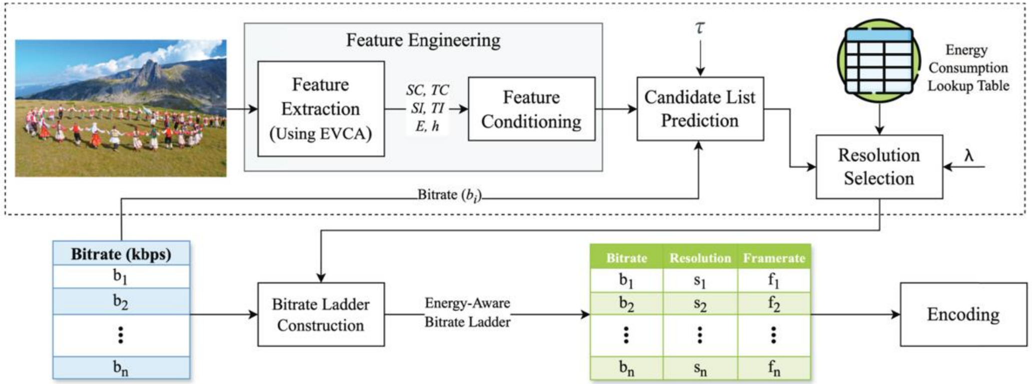


Fig. 1. The workflow of LiveESTR method.

THE PERFORMANCE COMPARISON OF LIVEESTR METHOD WITH A STATIC BITRATE LADDER AND GROUND TRUTH METHODS IN DIFFERENT THRESHOLDS

τ	BD-Rate (%)			BD-VMAF			BDDE (%)			BDEE (%)		
	Static ladder	Ground truth	LiveESTR	Static ladder	Ground truth	LiveESTR	Static ladder	Ground truth	LiveESTR	Static ladder	Ground truth	LiveESTR
1.0	50.4	-6.4	-0.9	-6.0	1.5	0.8	-30.7	-25.7	-29.4	-37.2	-16.1	-18.6
2.0	50.4	-2.2	2.1	-6.0	1.0	0.3	-30.7	-30.0	-29.7	-37.2	-22.5	-22.6
3.0	50.4	1.7	5.9	-6.0	0.4	-0.3	-30.7	-33.5	-32.6	-37.2	-27.8	-28.5
4.0	50.4	5.2	11.0	-6.0	-0.2	-1.1	-30.7	-36.0	-36.9	-37.2	-31.9	-33.7
5.0	50.4	9.8	16.0	-6.0	-0.9	-1.8	-30.7	-37.3	-38.2	-37.2	-34.1	-35.3

UNIVERSITÄT KLAGENFURT

ATHENA
Christian Doppler (CD) Laboratory

Bitrate Ladder Construction

Live Video Streaming

Table 5. Performance comparison in YPSNR / VMAF of the considered models on the proposed dataset with HEVC encoder. The top result for each encoder type is highlighted in boldface.

Encoder type		HEVC software encoding					
	Model \ Metric	R2 ↑	SROCC ↑	PLCC ↑	Accuracy ↑	BD-BR vs EEL ↓	BD-BR vs SL ↓
VoD-HandC	ExtraTrees	**0.6479 / 0.4929**	**0.6730 / 0.6286**	**0.8154 / 0.7131**	**0.8852 / 0.8685**	**1.776% / 2.545%**	**-5.997% / -5.583%**
	Random Forests	0.6079 / 0.4293	0.6452 / 0.6087	0.7847 / 0.6679	0.8814 / 0.8626	2.365% / 2.670%	-5.886% / -5.721%
	XGBoost	0.5659 / 0.3877	0.6312 / 0.5867	0.7767 / 0.6541	0.8775 / 0.8606	2.909% / 3.331%	-5.524% / -4.943%
	LightGBM	0.5487 / 0.4095	0.6436 / 0.5762	0.7518 / 0.6496	0.8665 / 0.8397	2.023% / 3.841%	-5.293% / -4.785%
Live-HandC	ExtraTrees	0.2700 / 0.3250	0.4941 / 0.5341	0.5411 / 0.5832	0.8354 / 0.8396	3.448% / 4.037%	-4.805% / -5.179%
	Random Forests	0.2627 / 0.2301	0.5083 / 0.5363	0.5303 / 0.5060	0.8329 / 0.8303	3.870% / 4.310%	-4.784% / -4.803%
	XGBoost	0.2507 / 0.1920	0.5181 / 0.4978	0.5317 / 0.4471	0.8303 / 0.8299	3.866% / 4.041%	-4.828% / -4.585%
	LightGBM	0.2438 / 0.1687	0.5172 / 0.4279	0.5088 / 0.4169	0.8315 / 0.8149	3.840% / 4.188%	-4.889% / -4.531%
Deep features	DensNet169	0.3306 / 0.3139	0.3778 / 0.5374	0.5703 / 0.5250	0.8793 / 0.8395	3.300% / 3.181%	-4.915% / -4.823%
	VGG16	0.2787 / 0.2511	0.3832 / 0.5531	0.5767 / 0.5083	0.8637 / 0.8227	3.378% / 3.329%	-4.845% / -4.751%
	ResNet-50	0.5652 / 0.3204	0.5057 / 0.5802	0.6349 / 0.5400	0.8805 / 0.8539	2.755% / 2.999%	-5.387% / -5.012%
	ConvNeXtBase	0.3188 / 0.2321	0.4788 / 0.4650	0.5884 / 0.5005	0.8756 / 0.8252	3.600% / 3.434%	-4.892% / -4.712%
Encoder type		HEVC hardware encoding					
	Model \ Metric	R2 ↑	SROCC ↑	PLCC ↑	Accuracy ↑	BD-BR vs EEL ↓	BD-BR vs SL ↓
VoD-HandC	ExtraTrees	**0.4728** / 0.3822	0.5927 / 0.4424	**0.7027** / 0.6452	**0.8373** / 0.7815	**2.644%** / 4.850%	-5.319% / -5.068%
	Random Forests	0.4113 / 0.3481	0.5475 / 0.3934	0.6527 / 0.6057	0.8278 / 0.7800	3.292% / 4.724%	-4.896% / -4.932%
	XGBoost	0.4139 / 0.2794	0.5219 / 0.3452	0.6536 / 0.5462	0.8228 / 0.7669	3.880% / 5.370%	-4.916% / -4.472%
	LightGBM	0.3039 / 0.1876	0.5304 / 0.3347	0.6428 / 0.4816	0.8029 / 0.7371	4.501% / 5.915%	-4.922% / -3.728%
Live-HandC	ExtraTrees	0.3188 / 0.2321	0.4788 / 0.4650	0.5884 / 0.5005	0.7894 / 0.7715	4.610% / 5.100%	-4.209% / -3.533%
	Random Forests	0.2854 / 0.1866	0.4879 / 0.3989	0.5612 / 0.4020	0.7841 / 0.7613	5.205% / 5.921%	-3.951% / -3.258%
	XGBoost	0.2490 / 0.1391	0.4843 / 0.3629	0.5293 / 0.4686	0.7793 / 0.7627	5.049% / 5.930%	-4.094% / -2.807%
	LightGBM	0.3039 / 0.1876	0.5304 / 0.3347	0.6428 / 0.4816	0.8029 / 0.7371	4.501% / 5.915%	-4.922% / -3.728%
Deep features	DensNet169	0.3750 / **0.4408**	0.5710 / **0.5477**	0.6109 / **0.7133**	0.8272 / **0.8143**	3.059% / 4.333%	-4.967% / -5.018%
	VGG16	0.3422 / 0.3559	0.5584 / 0.4899	0.6159 / 0.6239	0.7649 / 0.7529	3.750% / 5.335%	-4.807% / -3.931%
	ResNet-50	0.3466 / 0.4229	**0.5965** / 0.5251	0.6354 / 0.7025	0.8278 / 0.8003	2.811% / **4.150%**	**-5.139% / -5.533%**
	ConvNeXtBase	0.2440 / 0.3490	0.5013 / 0.5127	0.5284 / 0.6253	0.7420 / 0.7544	3.916% / 5.417%	-4.745% / -3.590%

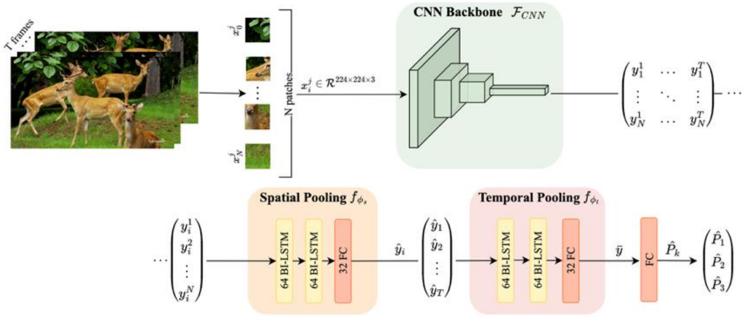


Fig. 6. The overall framework of the proposed DNN methods. The feature extraction module extracts spatial features $\boldsymbol{y}_i^j$ from patches $\boldsymbol{x}_i^j$. The spatial and temporal pooling modules aggregate features into a final vector $\bar{\boldsymbol{y}}$. Finally, the regression module uses the final vector $\bar{\boldsymbol{y}}$ to predict the cross-over bitrates $\hat{P}_k$.

UNIVERSITÄT KLAGENFURT

ATHENA
Christian Doppler (CD) Laboratory

Bitrate Ladder Construction

Live Video Streaming

Table 5. Performance comparison in YPSNR / VMAF of the considered models on the proposed dataset with HEVC encoder. The top result for each encoder type is highlighted in boldface.

Encoder type		HEVC software encoding					
	Model \ Metric	R2 ↑	SROCC ↑	PLCC ↑	Accuracy ↑	BD-BR vs EEL ↓	BD-BR vs SL ↓
VoD-HandC	ExtraTrees	**0.6479** / **0.4929**	**0.6730** / **0.6286**	**0.8154** / **0.7131**	**0.8852** / **0.8685**	**1.776%** / **2.545%**	**-5.997%** / **-5.583%**
	Random Forests	0.6079 / 0.4293	0.6452 / 0.6087	0.7847 / 0.6679	0.8814 / 0.8626	2.365% / 2.670%	-5.886% / -5.721%
	XGBoost	0.5659 / 0.3877	0.6312 / 0.5867	0.7767 / 0.6541	0.8775 / 0.8606	2.909% / 3.331%	-5.524% / -4.943%
	LightGBM	0.5487 / 0.4095	0.6436 / 0.5762	0.7518 / 0.6496	0.8665 / 0.8397	2.023% / 3.841%	-5.293% / -4.785%
Live-HandC	ExtraTrees	0.2700 / 0.3250	0.4941 / 0.5341	0.5411 / 0.5832	0.8354 / 0.8396	3.448% / 4.037%	-4.805% / -5.179%
	Random Forests	0.2627 / 0.2301	0.5083 / 0.5363	0.5303 / 0.5060	0.8329 / 0.8303	3.870% / 4.310%	-4.784% / -4.803%
	XGBoost	0.2507 / 0.1920	0.5181 / 0.4978	0.5317 / 0.4471	0.8303 / 0.8299	3.866% / 4.041%	-4.828% / -4.585%
	LightGBM	0.2438 / 0.1687	0.5172 / 0.4279	0.5088 / 0.4169	0.8315 / 0.8149	3.840% / 4.188%	-4.889% / -4.531%
Deep features	DensNet169	0.3306 / 0.3139	0.3778 / 0.5374	0.5703 / 0.5250	0.8793 / 0.8395	3.300% / 3.181%	-4.915% / -4.823%
	VGG16	0.2787 / 0.2511	0.3832 / 0.5531	0.5767 / 0.5083	0.8637 / 0.8227	3.378% / 3.329%	-4.845% / -4.751%
	ResNet-50	0.5652 / 0.3204	0.5057 / 0.5802	0.6349 / 0.5400	0.8805 / 0.8539	2.755% / 2.999%	-5.387% / -5.012%
	ConvNeXtBase	0.3188 / 0.2321	0.4788 / 0.4650	0.5884 / 0.5005	0.8756 / 0.8252	3.600% / 3.434%	-4.892% / -4.712%
Encoder type		HEVC hardware encoding					
	Model \ Metric	R2 ↑	SROCC ↑	PLCC ↑	Accuracy ↑	BD-BR vs EEL ↓	BD-BR vs SL ↓
VoD-HandC	ExtraTrees	**0.4728** / 0.3822	0.5927 / 0.4424	**0.7027** / 0.6452	**0.8373** / 0.7815	**2.644%** / 4.850%	-5.319% / -5.068%
	Random Forests	0.4113 / 0.3481	0.5475 / 0.3934	0.6527 / 0.6057	0.8278 / 0.7800	3.292% / 4.724%	-4.896% / -4.932%
	XGBoost	0.4139 / 0.2794	0.5219 / 0.3452	0.6536 / 0.5462	0.8228 / 0.7669	3.880% / 5.370%	-4.916% / -4.472%
	LightGBM	0.3039 / 0.1876	0.5304 / 0.3347	0.6428 / 0.4816	0.8029 / 0.7371	4.501% / 5.915%	-4.922% / -3.728%
Live-HandC	ExtraTrees	0.3188 / 0.2321	0.4788 / 0.4650	0.5884 / 0.5005	0.7894 / 0.7715	4.610% / 5.100%	-4.209% / -3.533%
	Random Forests	0.2854 / 0.1866	0.4879 / 0.3989	0.5612 / 0.4020	0.7841 / 0.7613	5.205% / 5.921%	-3.951% / -3.258%
	XGBoost	0.2490 / 0.1391	0.4843 / 0.3629	0.5293 / 0.4686	0.7793 / 0.7627	5.049% / 5.930%	-4.094% / -2.807%
	LightGBM	0.3039 / 0.1876	0.5304 / 0.3347	0.6428 / 0.4816	0.8029 / 0.7371	4.501% / 5.915%	-4.922% / -3.728%
Deep features	DensNet169	0.3750 / **0.4408**	0.5710 / **0.5477**	0.6109 / **0.7133**	0.8272 / **0.8143**	3.059% / 4.333%	-4.967% / -5.018%
	VGG16	0.3422 / 0.3559	0.5584 / 0.4899	0.6159 / 0.6239	0.7649 / 0.7529	3.750% / 5.335%	-4.807% / -3.931%
	ResNet-50	0.3466 / 0.4229	**0.5965** / 0.5251	0.6354 / 0.7025	0.8278 / 0.8003	2.811% / **4.150%**	**-5.139%** / **-5.533%**
	ConvNeXtBase	0.2440 / 0.3490	0.5013 / 0.5127	0.5284 / 0.6253	0.7420 / 0.7544	3.916% / 5.417%	-4.745% / -3.590%

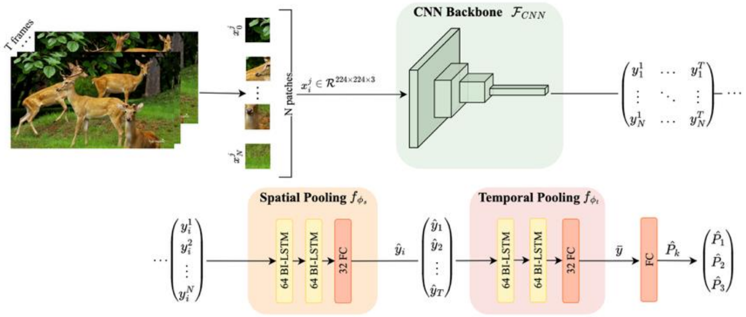
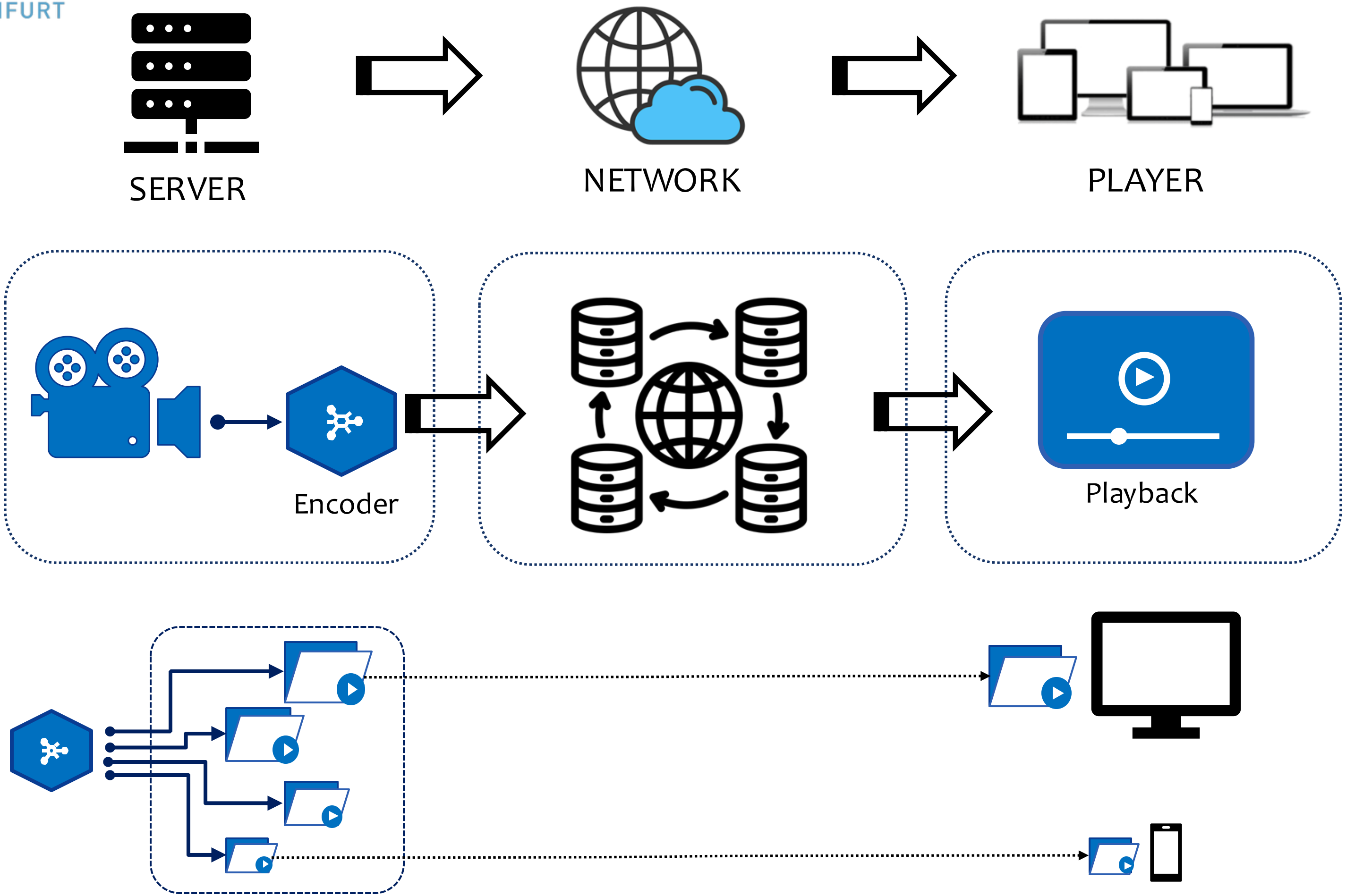
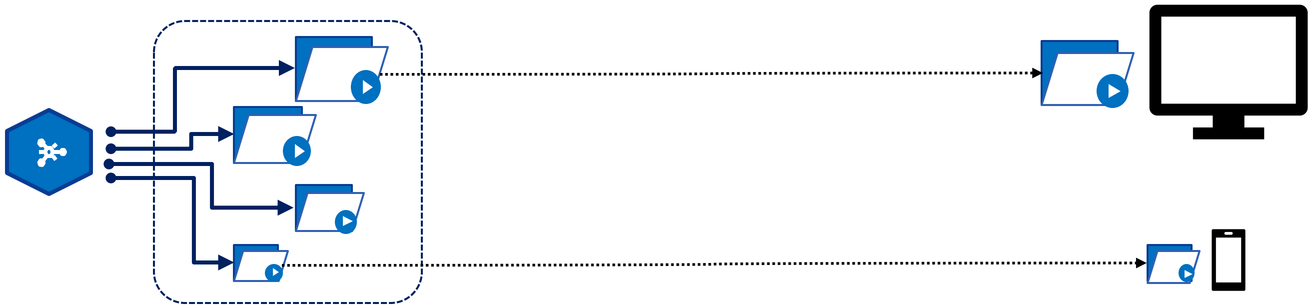
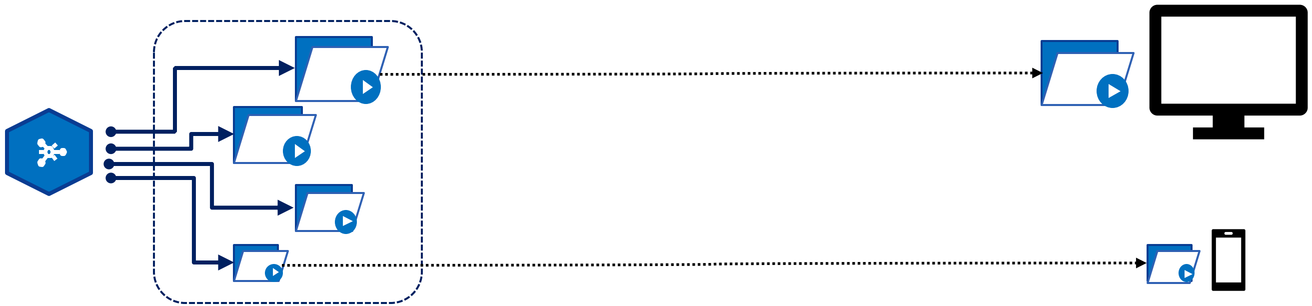


Fig. 6. The overall framework of the proposed DNN methods. The feature extraction module extracts spatial features $\boldsymbol{y}_i^j$ from patches $\boldsymbol{x}_i^j$. The spatial and temporal pooling modules aggregate features into a final vector $\bar{\boldsymbol{y}}$. Finally, the regression module uses the final vector $\bar{\boldsymbol{y}}$ to predict the cross-over bitrates $\hat{P}_k$.

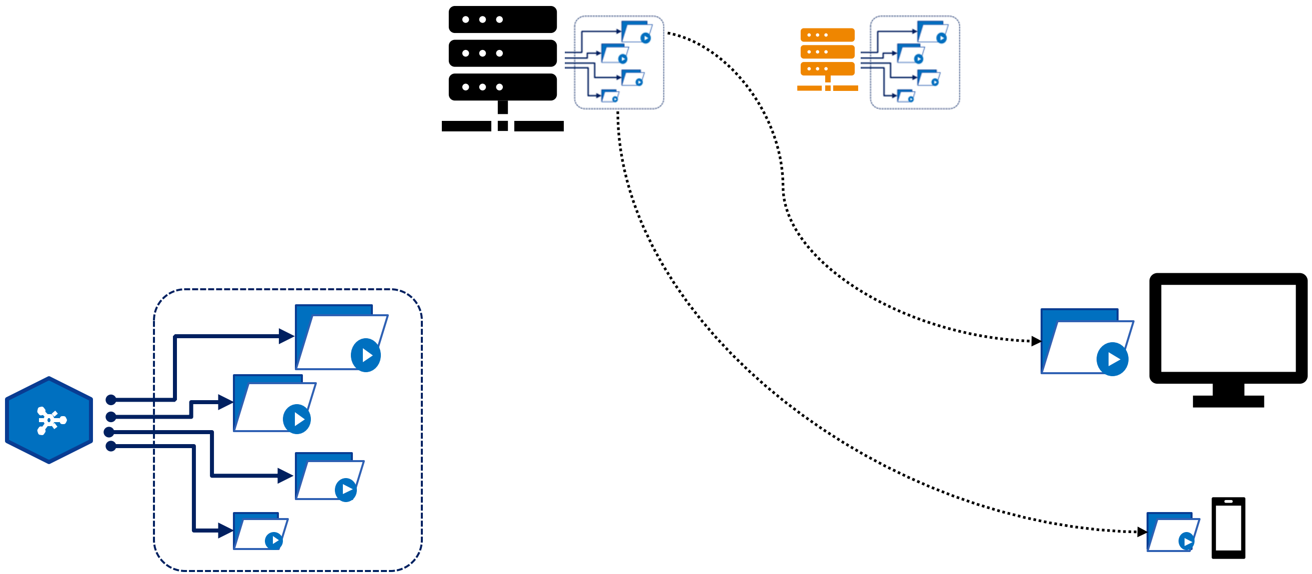

SERVER
NETWORK
PLAYER
Encoder
Playback


SERVER
NETWORK
PLAYER



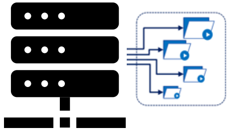


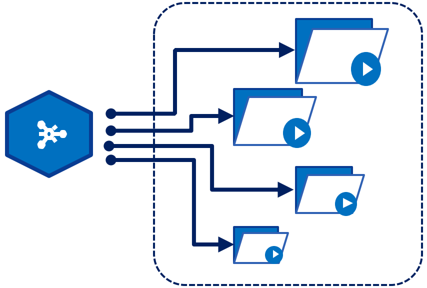
LwTE: Light-weight Transcoding at the Edge, 2021

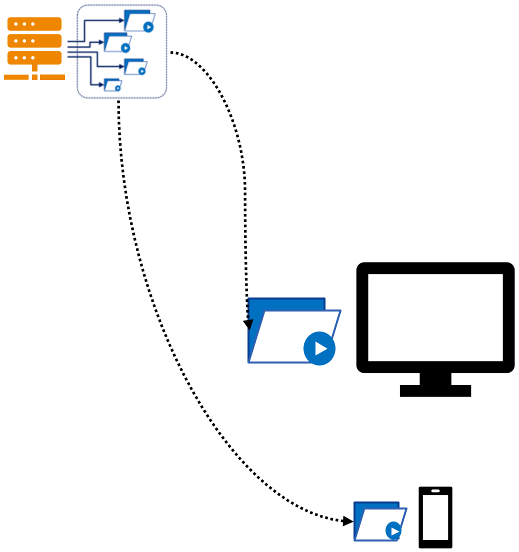


LwTE: Light-weight Transcoding at the Edge, 2021

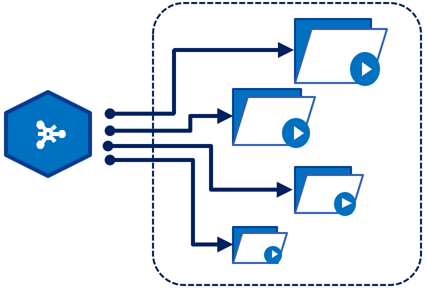

SERVER
NETWORK
PLAYER

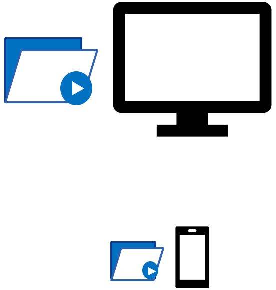


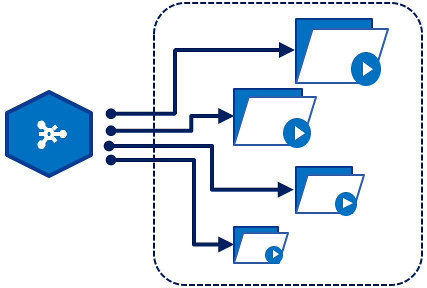


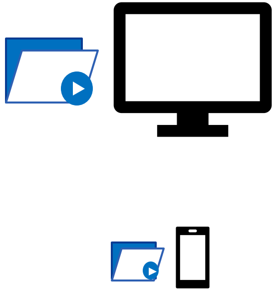



SERVER
NETWORK
PLAYER

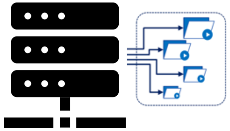


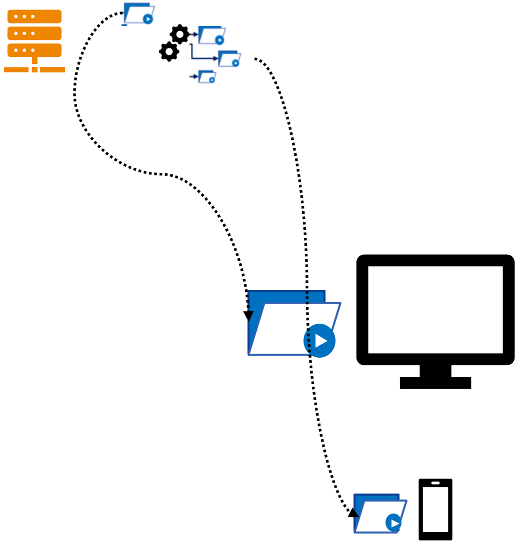


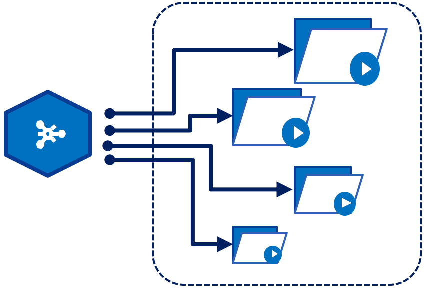




LwTE: Light-weight Transcoding at the Edge, 2021



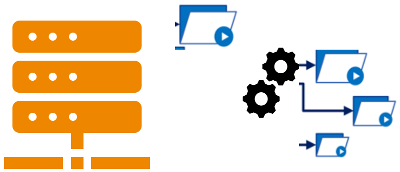


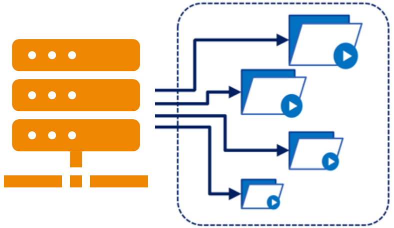


LwTE: Light-weight Transcoding at the Edge, 2021

Light-weight Transcoding

- Transcoding is expensive:
  - Optimal Encoding Decision
  - Quantization and Entropy Encoding





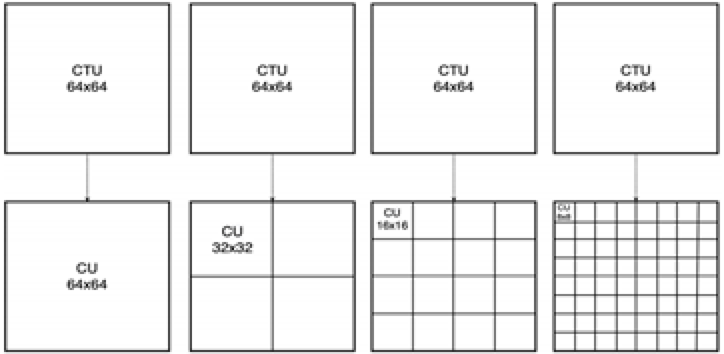
LwTE: Light-weight Transcoding at the Edge, 2021

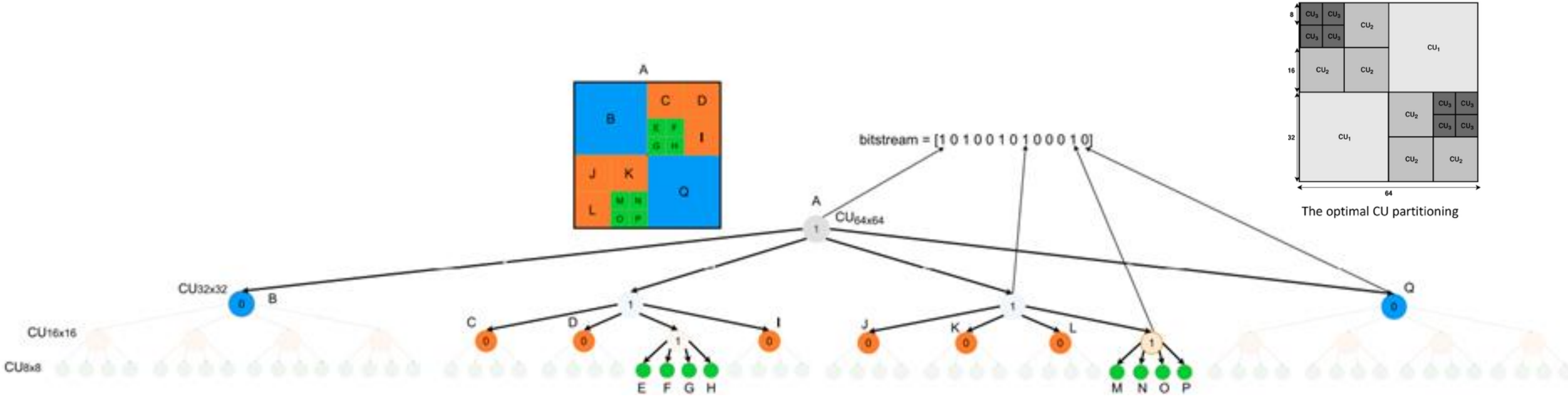
UNIVERSITÄT KLAGENFURT

ATHENA Christian Doppler (CD) Laboratory

Light-weight Transcoding

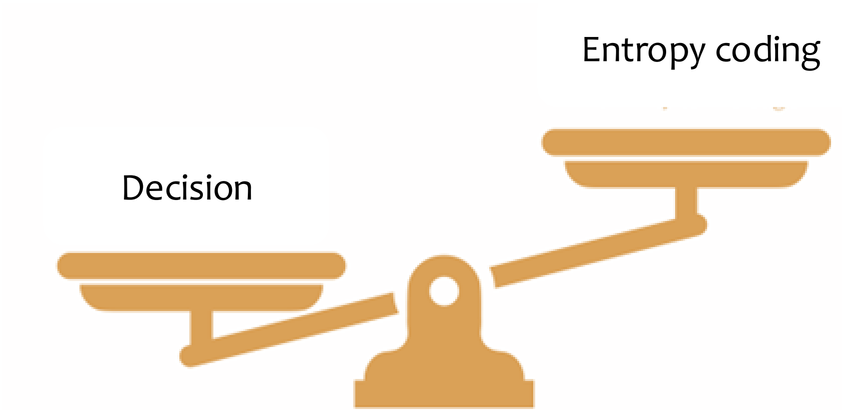
- Transcoding is expensive:
  - Optimal Encoding Decision
  - Quantization and Entropy Encoding



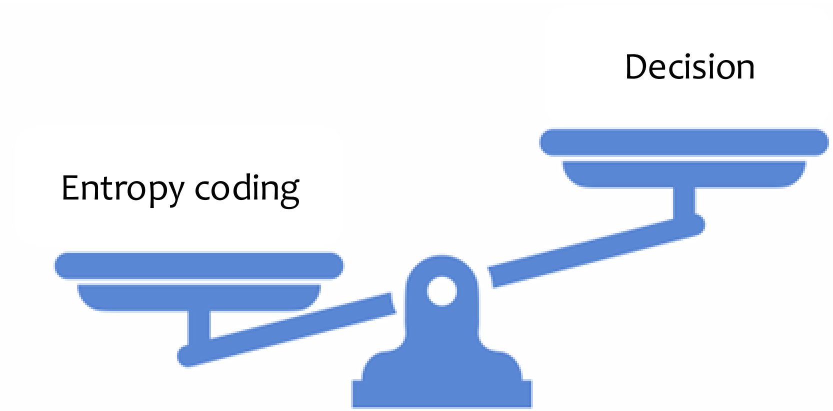


Light-weight Transcoding

- Transcoding is expensive:
  - Optimal Encoding Decision
  - Quantization and Entropy coding



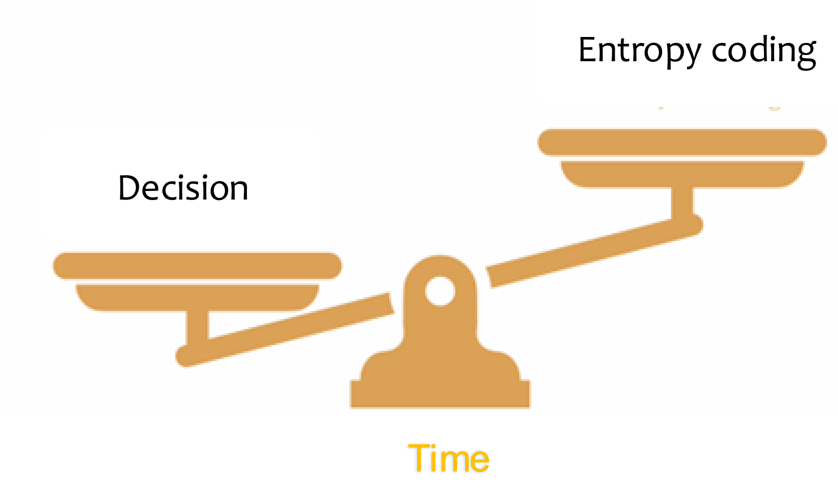
Time

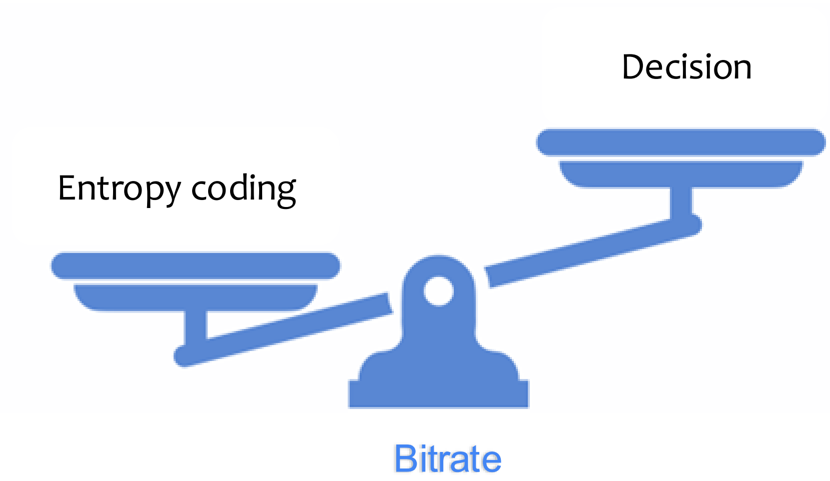


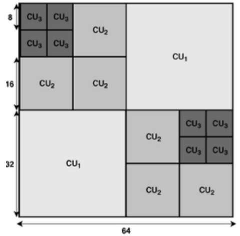
Bitrate

Light-weight Transcoding

- Finding the optimal CTU partitioning takes the majority of the encoding time
- Signaling the optimal CTU partitioning in bitsream requires minimal bits

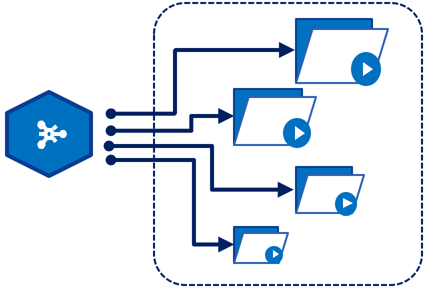


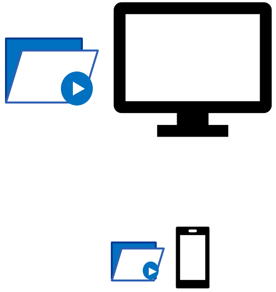




The optimal CU partitioning

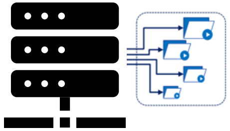
- The optimal CTU partitioning are stored as metadata in Edge servers
- During the transcoding, they are used to avoid the brute-force search process

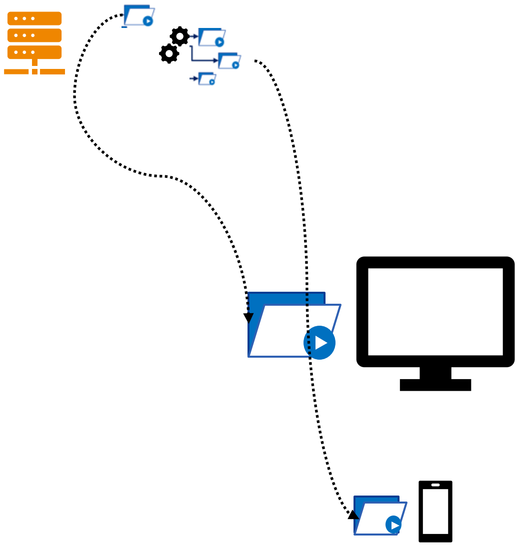


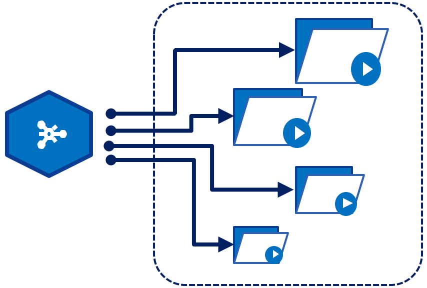


UNIVERSITÄT KLAGENFURT

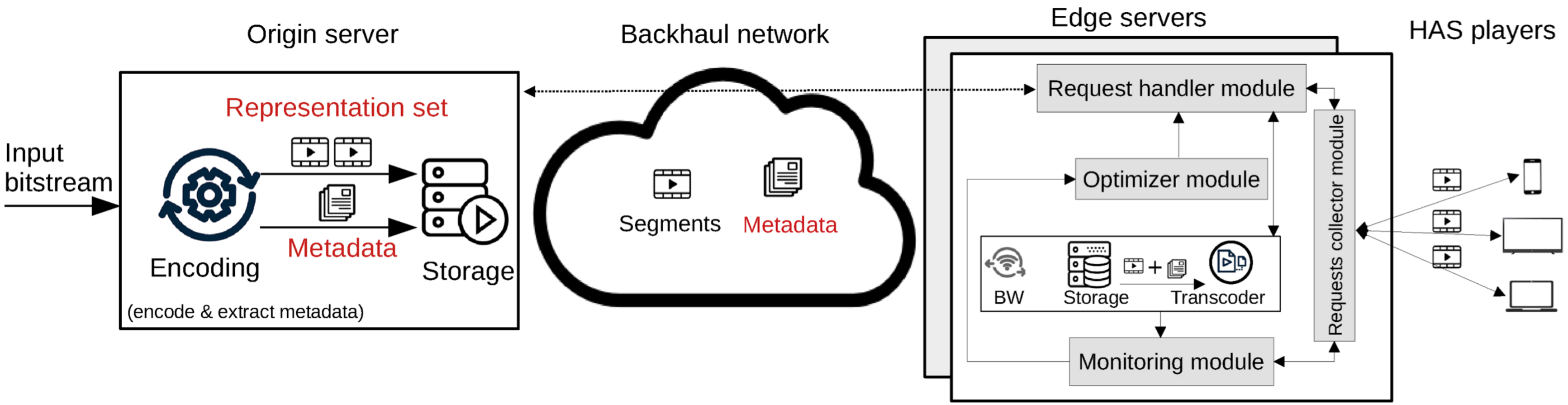
ATHENA
Christian Doppler (CD) Laboratory





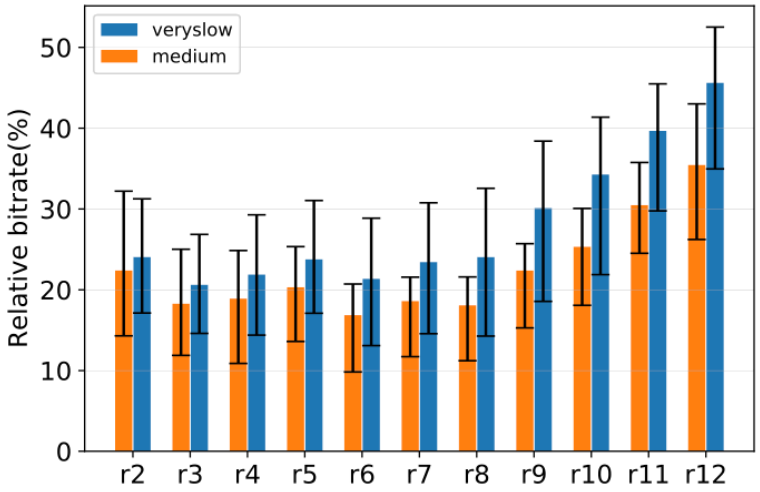


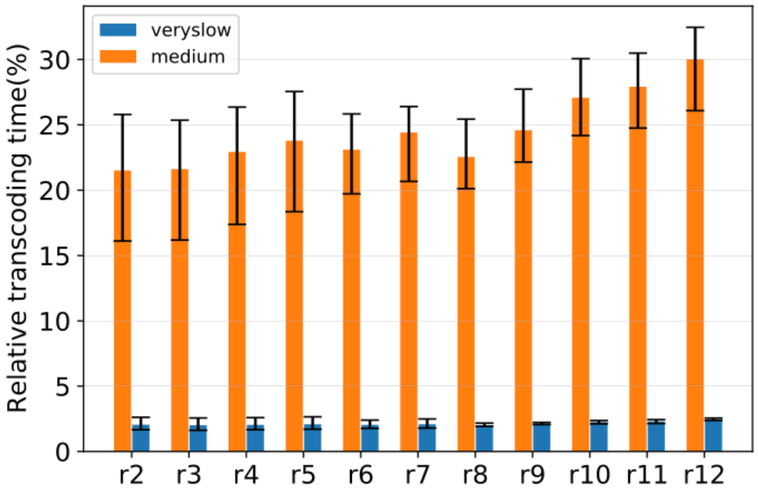
Light-weight Transcoding



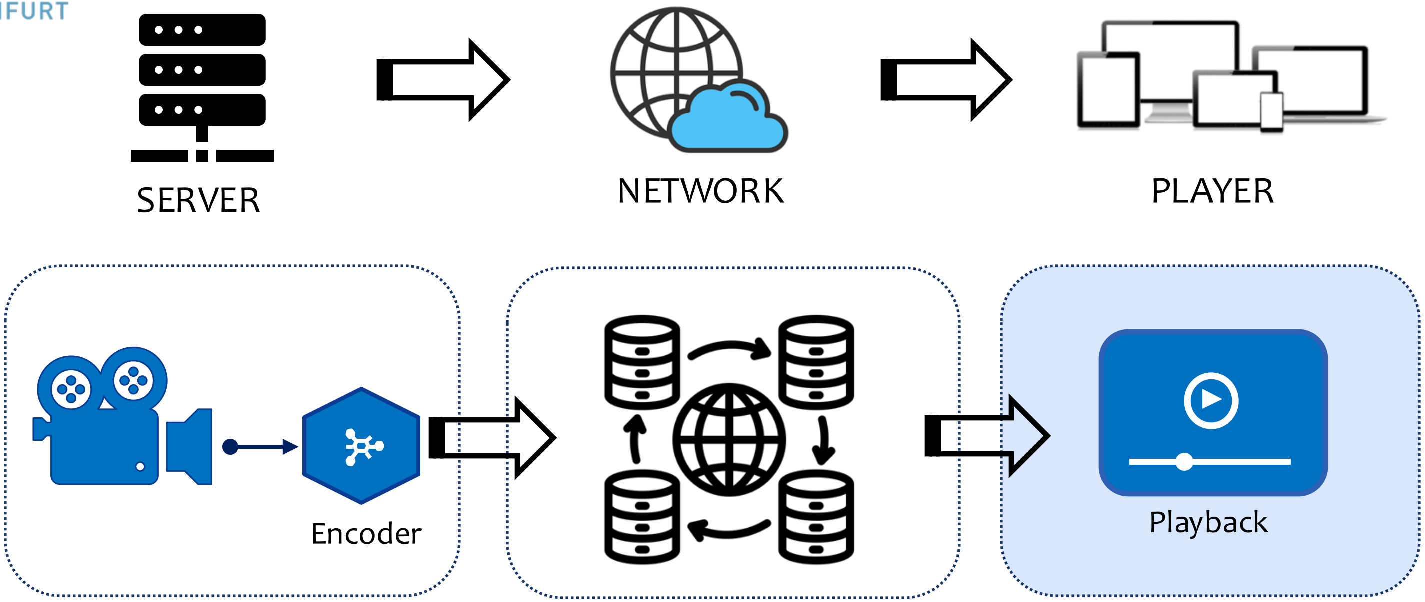
CD-LwTE architecture.

Light-weight Transcoding



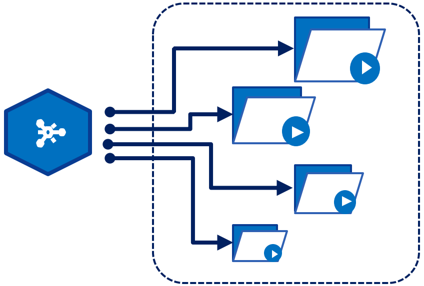


LwTE: Light-weight Transcoding at the Edge, 2021


SERVER
NETWORK
PLAYER
Encoder
Playback

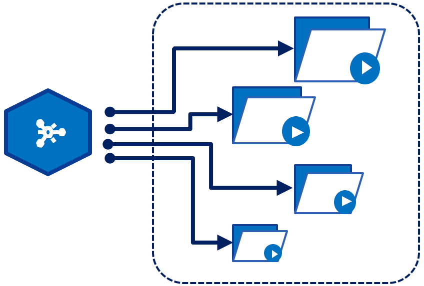
Quality of Experience

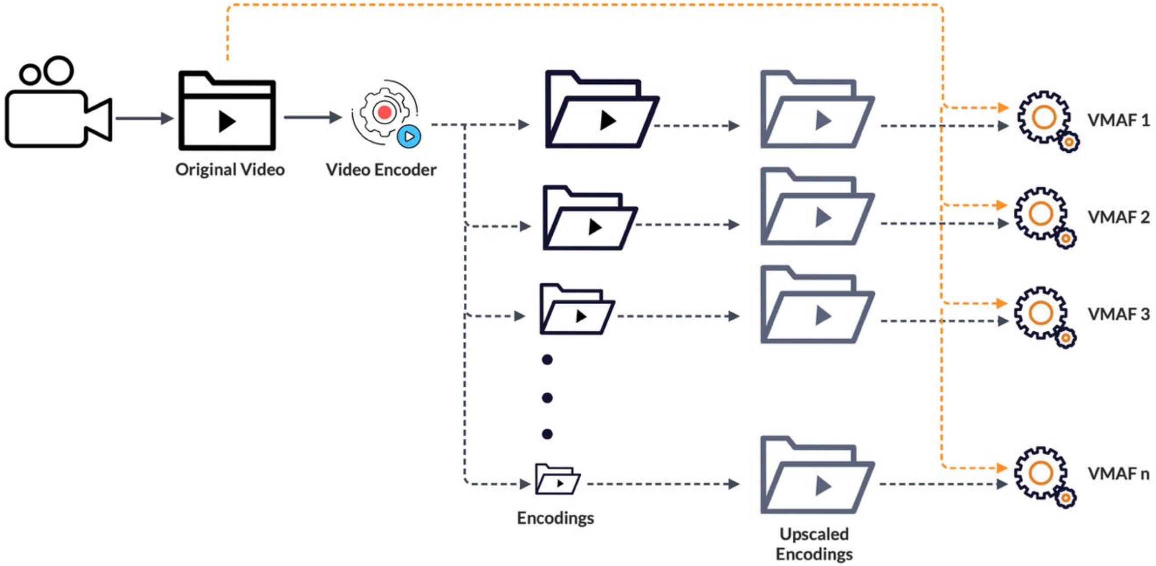
VQM4HAS: A Real-time Video Quality Metric for HAS



Quality of Experience

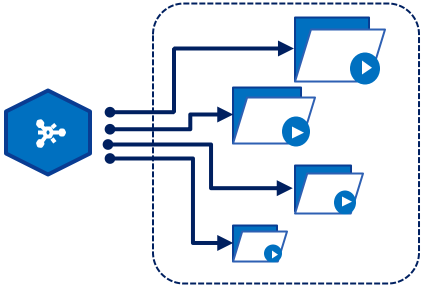
VQM4HAS: A Real-time Video Quality Metric for HAS

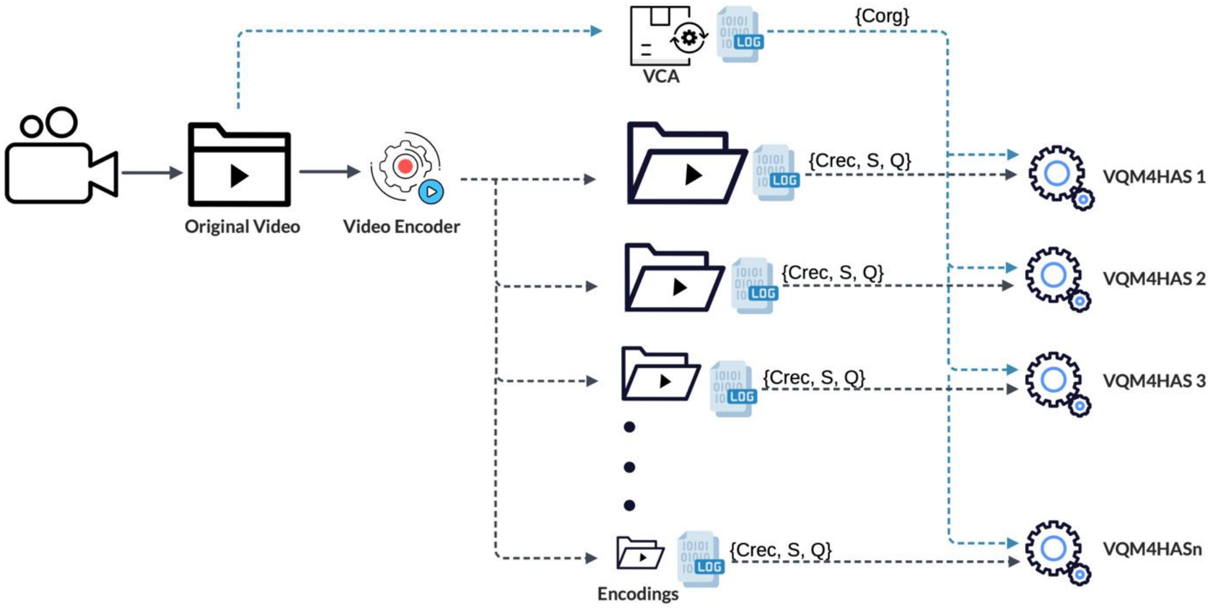




Quality of Experience

VQM4HAS: A Real-time Video Quality Metric for HAS





Quality of Experience

VQM4HAS: A Real-time Video Quality Metric for HAS

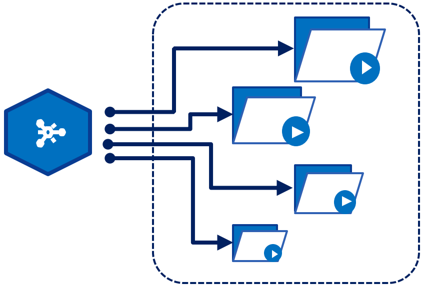
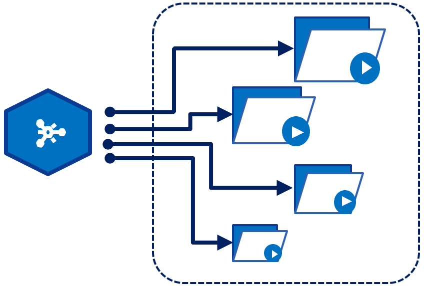


TABLE I: Description of encoding statistics when *–csv-log-level* is greater than or equal to 1.

Parameter	**Description**	**Used for evaluation**	**Name**
Encode Order	The frame order in which the encoder encodes.	No	
Type	Slice type of the frame.	No	
POC	Picture Order Count - The display order of the frames.	No	
QP	Quantization Parameter decided for the frame.	Yes	*qp*
Bits	Number of bits consumed by the frame.	Yes	*bit*
Scenecut	1 if the frame is a scenecut, 0 otherwise.	No	
RateFactor	Applicable only when CRF is enabled.	No	
BufferFill	Bits available for the next frame. Includes bits carried over from the current frame.	No	
BufferFillFinal	Buffer bits available after removing the frame out of CPB.	No	
UnclippedBufferFillFinal	Unclipped buffer bits available after removing the frame out of CPB only used for csv logging purposes.	No	
Latency	Latency in terms of number of frames between when the frame was given in and when the frame is given out.	No	
Ref lists	POC of references in lists 0 and 1 for the frame.	No	

Quality of Experience

VQM4HAS: A Real-time Video Quality Metric for HAS



Parameter	Description	Used for evaluation	Name
	Analysis statistics		
CU Statistics	Percentage of CU modes.	No	
Distortion	Average luma and chroma distortion.	Yes	$luma\_dist$, $chroma\_dist$
Psy Energy	Average psy energy calculated as the sum of absolute difference between source and recon energy.	Yes	psy_energy
Residual Energy	Average residual energy.	Yes	res_energy
Luma/Chroma Values	minimum, maximum and average luma and chroma values of source for each frame.	Yes	$luma\_avg$, cr_avg,cb_avg
PU Statistics	Percentage of PU modes at each depth.	No	
	Performance statistics		
DecideWait ms	number of milliseconds the frame encoder had to wait.	No	
Row0Wait ms	number of milliseconds since the frame encoder received a frame to encode before its first row of CTUs is allowed to begin compression.	No	
Wall time ms	number of milliseconds between the first CTU being ready to be compressed and the entire frame being compressed and the output NALs being completed.	No	
Ref Wait Wall ms	number of milliseconds between the first reference row being available and the last reference row becoming available.	No	
Total CTU time ms	the total time (measured in milliseconds) spent by worker threads compressing and filtering CTUs for this frame.	Yes	cpu_time
Stall Time ms	the number of milliseconds of the reported wall time that were spent with zero worker threads, aka all compression was completely stalled.	No	
Total frame time	Total time spent to encode the frame.	Yes	$frame\_time$
Avg WPP	the average number of worker threads working on this frame, at any given time.	No	
Row Blocks	the number of times a worker thread had to abandon the row of CTUs it was encoding.	No	

Quality of Experience

VQM4HAS: A Real-time Video Quality Metric for HAS

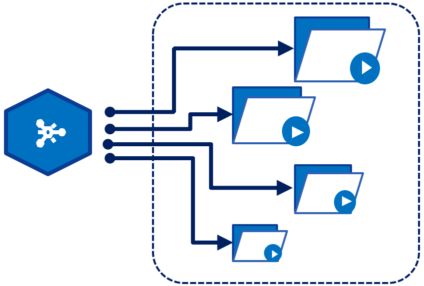
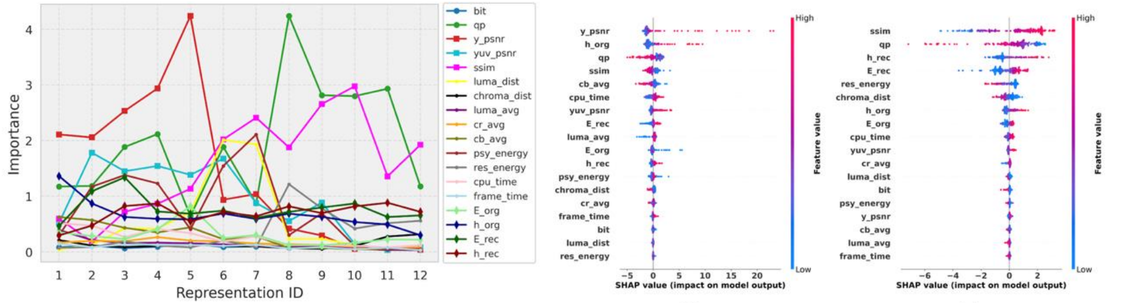


TABLE IV: PCC for *VQM4HAS* when predicting per-segment VMAF scores.

	Representation ID	1	2	3	4	5	6	7	8	9	10	11	12
Model	**Linear**	0.83	0.86	0.90	0.91	0.93	0.95	0.95	0.95	0.95	0.94	0.92	0.90
	Random Forest	0.87	0.90	0.94	0.94	0.95	0.95	0.95	0.96	0.95	0.95	0.95	0.94

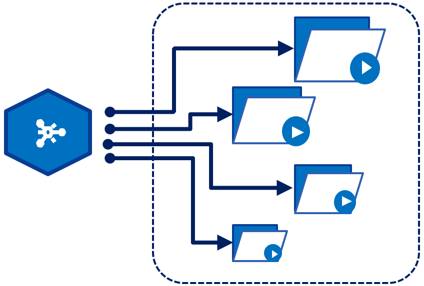


(a) (b) (c)

Fig. 4: (a) The feature importance of the features used to predict VMAF scores for all representations. SHAP summary plot for (b) the lowest and (c) highest bitrate representations.

Quality of Experience

VQM4HAS: A Real-time Video Quality Metric for HAS



	Resolution	540p	1080p	2160p
Method	**PSNR**	0.32	0.56	0.51
	SSIM	0.46	0.68	0.66
	VMAF	0.73	0.89	0.83
	P.1204.3	0.94	0.94	0.90
	VQM4HAS	0.96	0.93	0.93

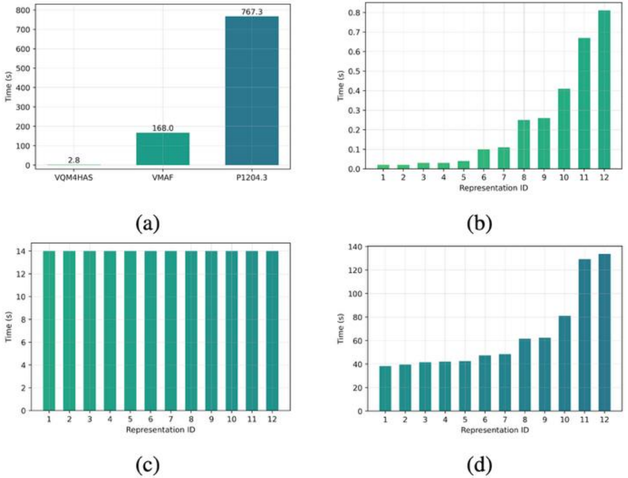
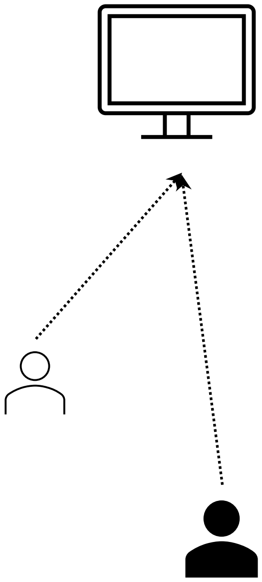
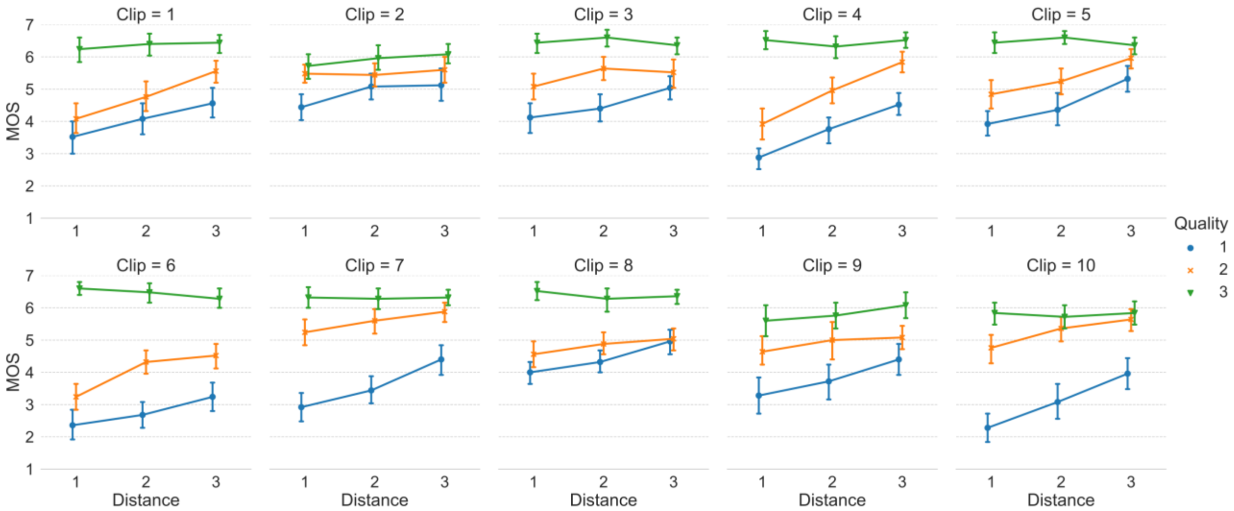


Fig. 9: (a) The time complexity of different methods when computing 12 representations of a 5-second video encoded with the HLS ladder parameters. The complexity breakdown across representations for (b) *VQM4HAS*, (c) VMAF, and (d) P.1204.3.

Quality of Experience

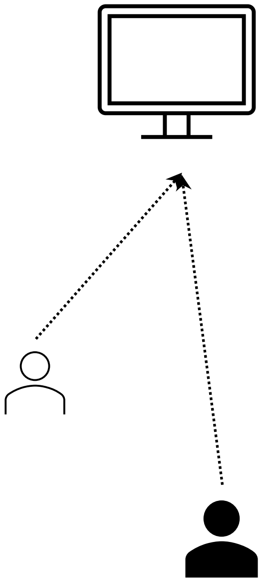
Impact of Viewing Distance [**In-Lab** Subjective Test]





Quality of Experience

Impact of Viewing Distance [**In-Lab** Subjective Test]



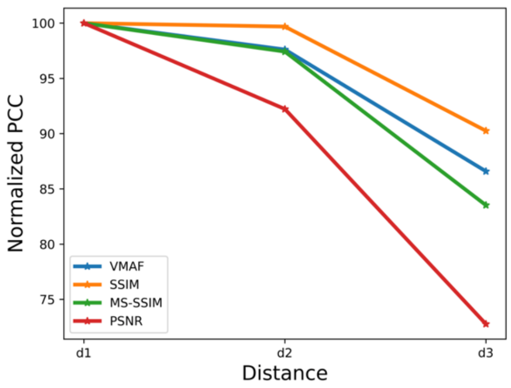
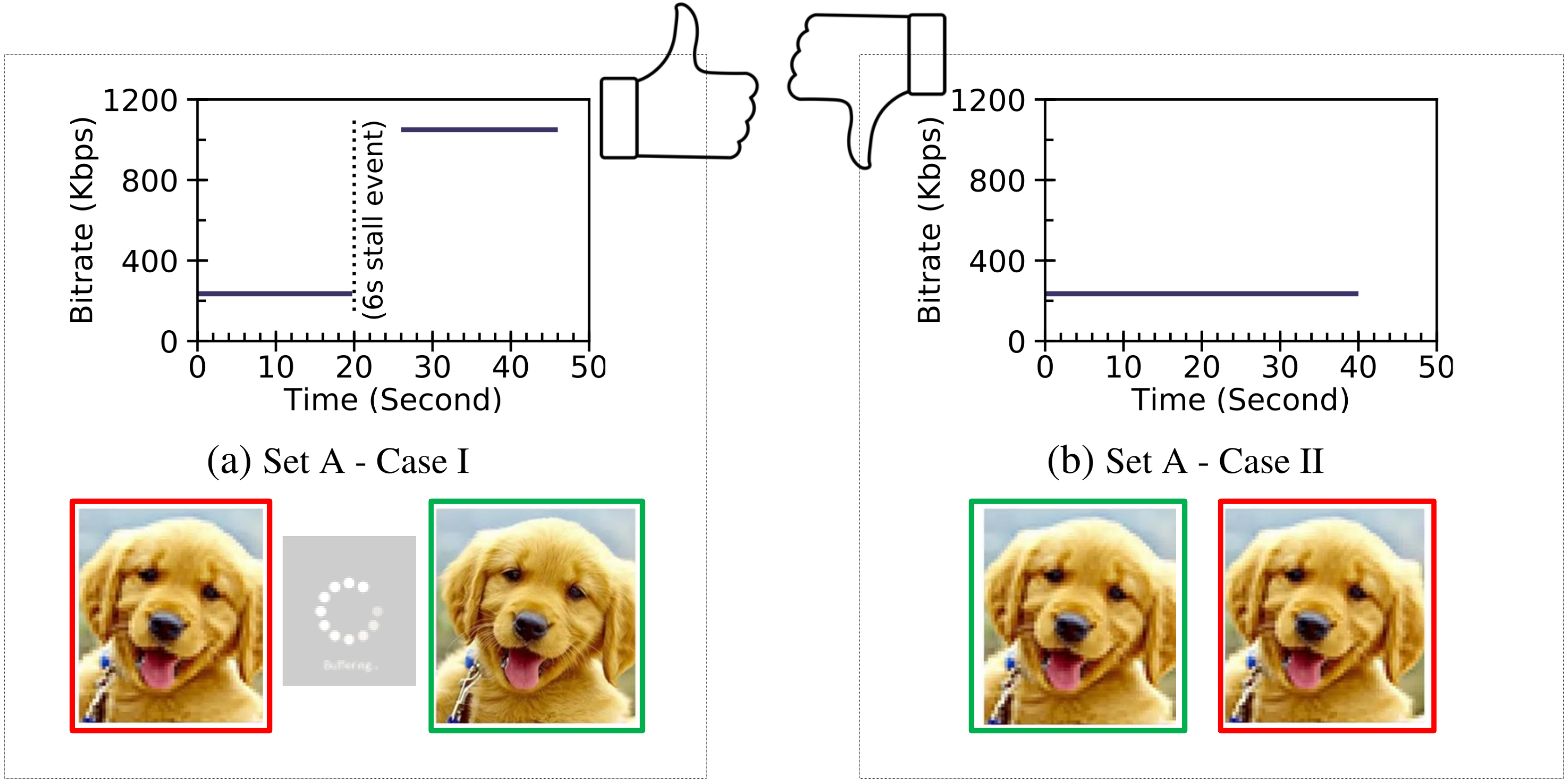


Fig. 5: Normalized Pearson correlation with subjective ratings for different objective quality metrics as function of distance (PCC at d1= 100%).

What Matters to Human Eye?

INTENSE: In-Depth Studies on Stall Events and Quality Switches and Their Impact on the Quality of Experience in HTTP Adaptive Streaming

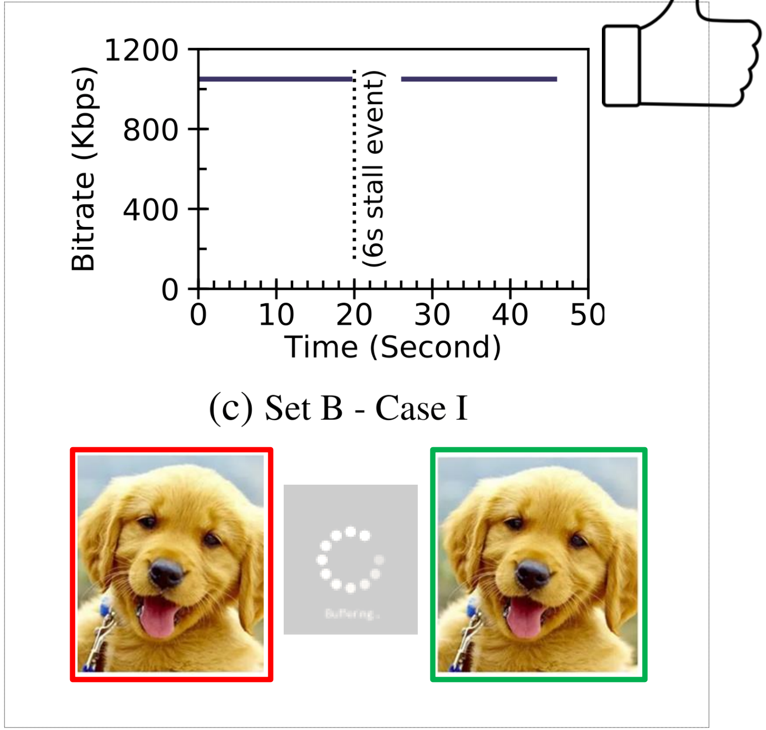
BABAK TARAGHI, (Member, IEEE), MINH NGUYEN, (Member, IEEE), HADI AMIRPOUR, (Member, IEEE), AND CHRISTIAN TIMMERER, (Senior Member, IEEE)
Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria



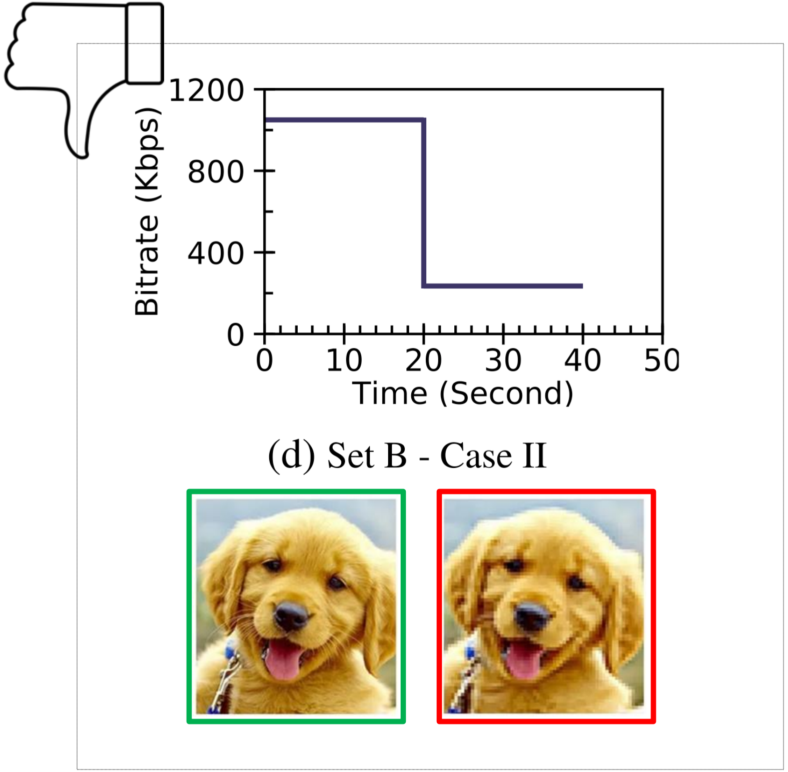
What Matters to Human Eye?

INTENSE: In-Depth Studies on Stall Events and Quality Switches and Their Impact on the Quality of Experience in HTTP Adaptive Streaming

BABAK TARAGHI, (Member, IEEE), MINH NGUYEN, (Member, IEEE), HADI AMIRPOUR, (Member, IEEE), AND CHRISTIAN TIMMERER, (Senior Member, IEEE)
Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria



(c) Set B - Case I



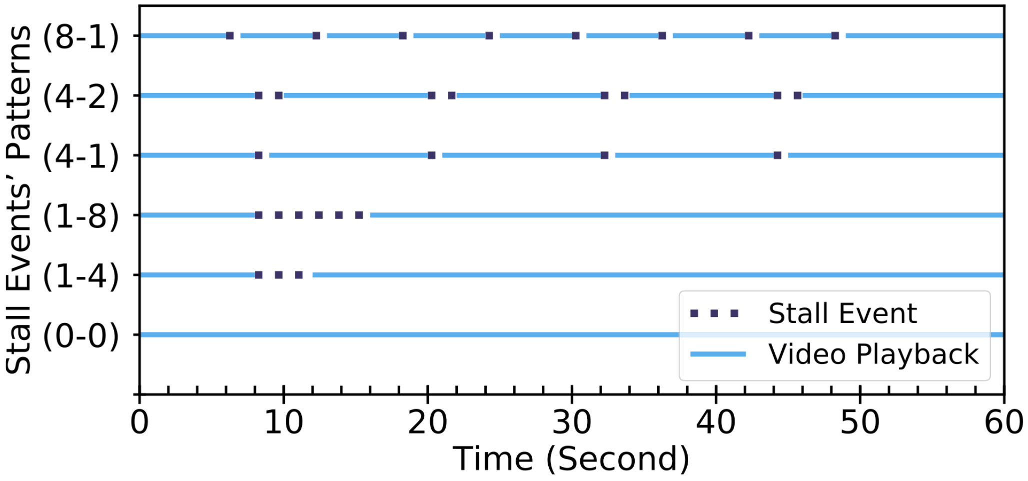
(d) Set B - Case II

INTENSE: In-Depth Studies on Stall Events and Quality Switches and Their Impact on the Quality of Experience in HTTP Adaptive Streaming

BABAK TARAGHI, (Member, IEEE), MINH NGUYEN, (Member, IEEE),
HADI AMIRPOUR, (Member, IEEE), AND CHRISTIAN TIMMERER, (Senior Member, IEEE)
Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria

What Matters to Human Eye?

- The analysis results demonstrate a preference for a longer stall event over stall events with high frequency but with the same total duration as the longer stall event.



UNIVERSITÄT KLAGENFURT

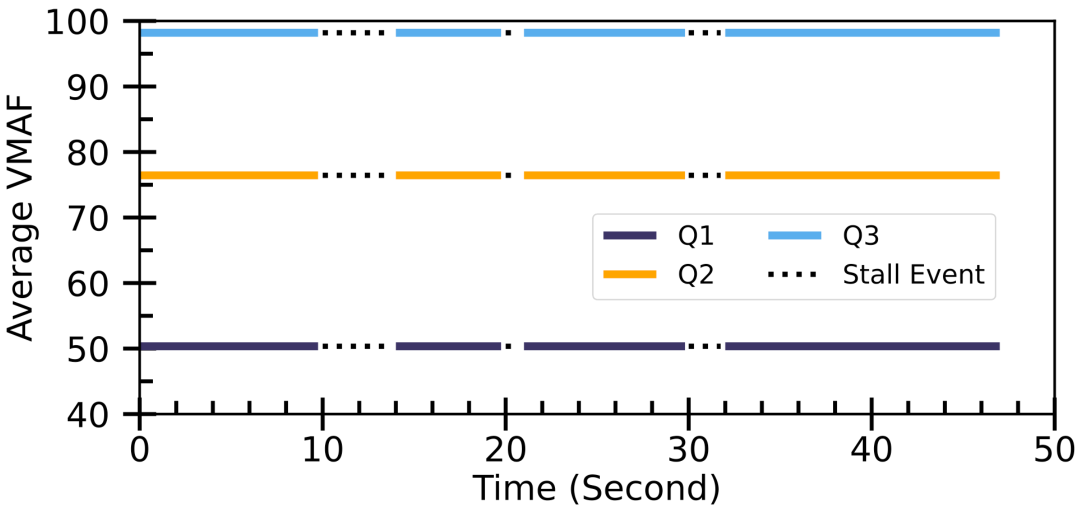
ATHENA
ian Doppler (CD) Laboratory

INTENSE: In-Depth Studies on Stall Events and Quality Switches and Their Impact on the Quality of Experience in HTTP Adaptive Streaming

BABAK TARAGHI, (Member, IEEE), MINH NGUYEN, (Member, IEEE), HADI AMIRPOUR, (Member, IEEE), AND CHRISTIAN TIMMERER, (Senior Member, IEEE)
Christian Doppler Laboratory ATHENA, Alpen-Adria-Universität Klagenfurt, 9020 Klagenfurt am Wörthersee, Austria

What Matters to Human Eye?

- It can be seen that stall events have a minor penalty on the QoE when the quality of videos is low.
- for the middle and high-quality videos, the stall event occurrence has a higher penalty on the perceived QoE than the same stall event at a low-quality video.



Current & Future Work + Open Challenges

- HTTP adaptive streaming: DASH, HLS, CMAF ✅
- Video coding for HAS (bitrate ladder optimizations) has been extensively researched, some niche problems seeking for solutions ⚠️
- Media over QUIC needs more attention in the multimedia research community ‼️
- New (immersive) modalities & coding formats in its infancy wrt streaming ‼️
- (Generative) AI for video streaming ‼️

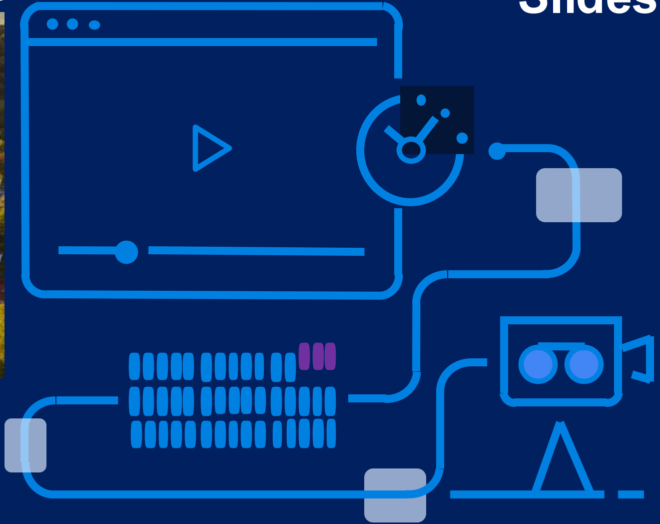
UNIVERSITÄT KLAGENFURT

ATHENA
Christian Doppler (CD) Laboratory

https://athena.itec.aau.at/

Slides

Video Coding Advancements in HTTP Adaptive Streaming

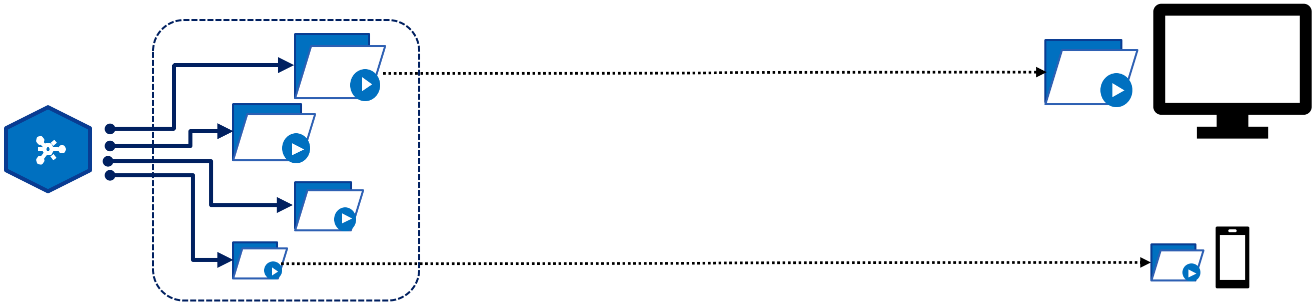


IEEE ICME 2025, June 30, 2025

Hadi Amirpour
Christian Timmerer

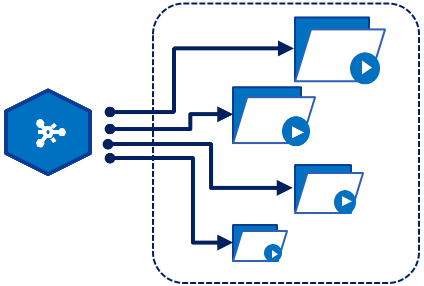
Alpen-Adria Universität Klagenfurt, Austria
Christian Doppler Laboratory ATHENA

Video Coding for HAS



Video Coding for HAS

Fast Multi-rate Encoding



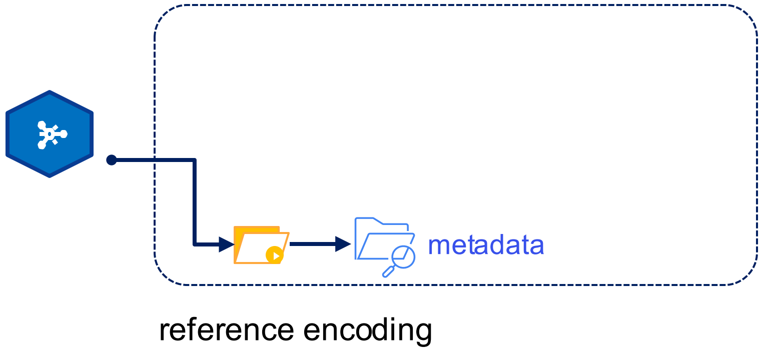
The same video

Encoded independently

Encoding decisions can be reused

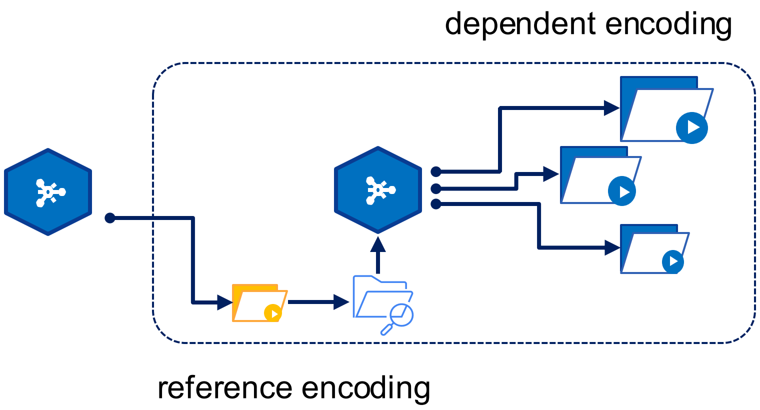
Video Coding for HAS

Fast Multi-rate Encoding



Video Coding for HAS

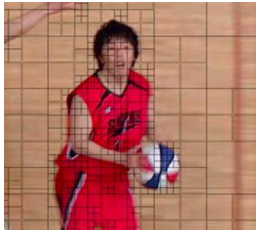
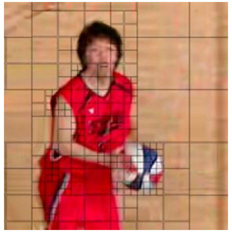
Fast Multi-rate Encoding



Video Coding for HAS

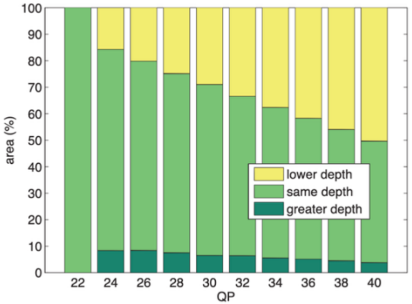
Fast Multi-rate Encoding

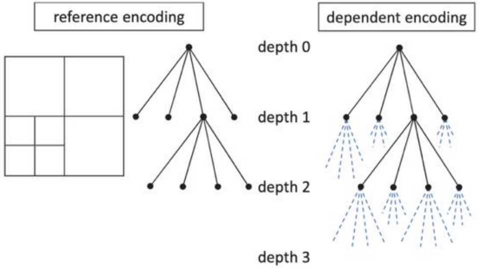


Video Coding for HAS

Fast Multi-rate Encoding





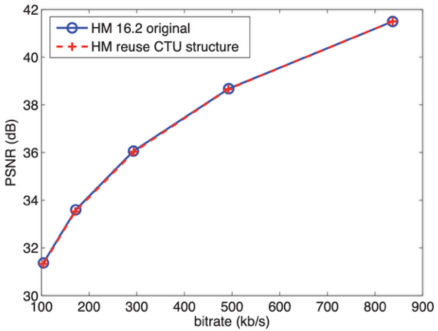
Efficient Multi-Rate Video Encoding for HEVC-Based Adaptive HTTP Streaming, TCSVT 2016

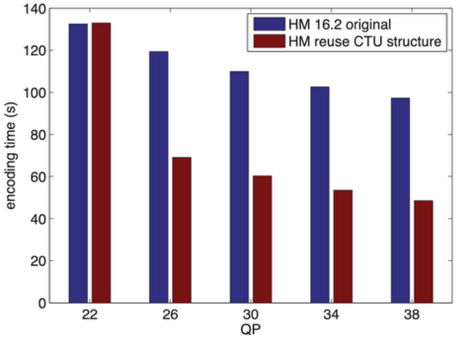
UNIVERSITÄT KLAGENFURT

ATHENA
Christian Doppler (CD) Laboratory

Video Coding for HAS

Fast Multi-rate Encoding





Video Coding for HAS

Fast Multi-rate Encoding

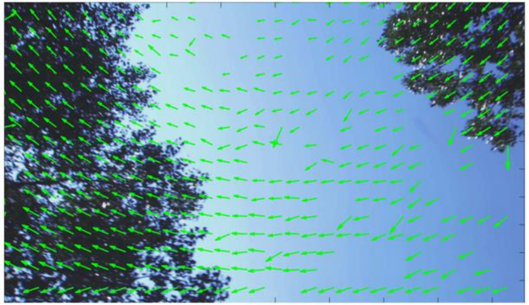


Table 1. Comparison of encoding results

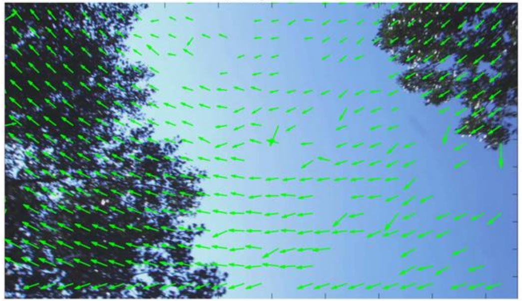
Sequence	BD-rate	BD-PSNR	ΔT
PeopleStreet (2560×1600)	0.85%	-0.04%	-19.02%
Kimono (1920×1080)	0.75%	-0.02%	-43.10%
ParkScene (1920×1080)	0.43%	-0.01%	-35.79%
BQMall (832×480)	0.57%	-0.02%	-28.60%
PartyScene (832×480)	0.26%	-0.01%	-22.10%
BasketballPass (416×240)	0.62%	-0.03%	-35.12%
BlowingBubbles (416×240)	0.43%	-0.02%	-20.03%
RaceHorses (416×240)	0.55%	-0.03%	-12.67%
Average	**0.56%**	**-0.02%**	**-27.05%**

Video Coding for HAS

Fast Multi-rate Encoding



(a) QP 22

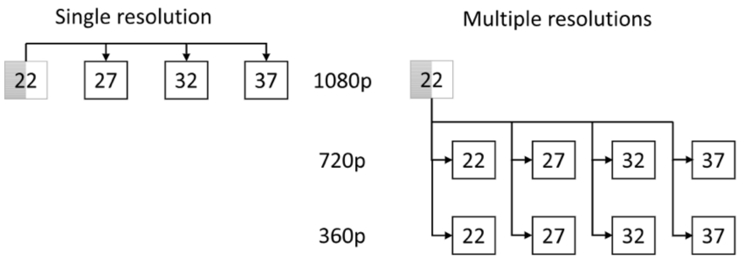


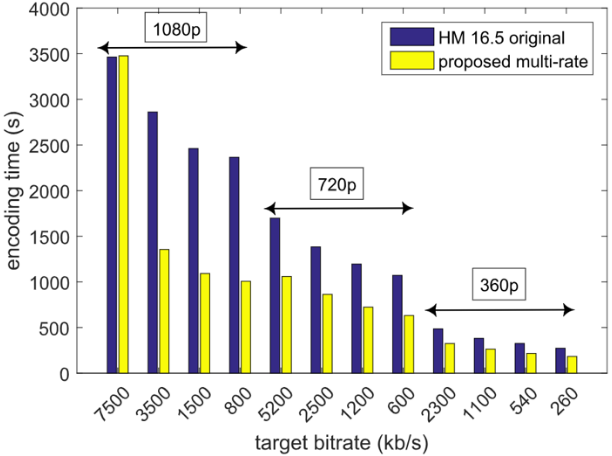
(b) QP 24

Sequence	BD-rate	BD-PSNR	ΔT
BlueSky	0.07%	−0.003 dB	−4.17%
CrowdRun	−0.08%	0.003 dB	−4.89%
DucksTakeOff	−0.09%	0.002 dB	−5.60%
Kimono	−0.06%	0.002 dB	−6.25%
ParkJoy	−0.20%	0.008 dB	−5.71%
ParkScene	−0.05%	0.002 dB	−2.73%
PedestrianArea	−0.19%	0.006 dB	−8.75%
Riverbed	−0.05%	0.002 dB	−11.58%
RushHour	−0.09%	0.002 dB	−6.16%
Sunflower	−0.52%	0.016 dB	−6.25%
Average	**−0.12%**	**0.004 dB**	**−6.21%**

Video Coding for HAS

Fast Multi-rate Encoding





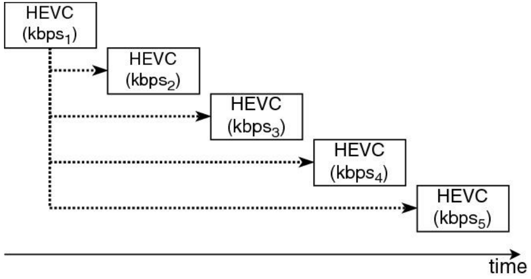
Efficient Multi-Rate Video Encoding for HEVC-Based Adaptive HTTP Streaming, TCSVT 2016

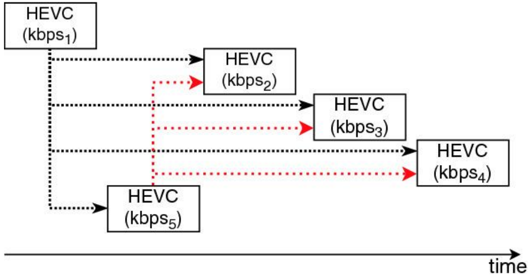
Video Coding for HAS

Fast Multi-rate Encoding

Video Coding for HAS

Fast Multi-rate Encoding





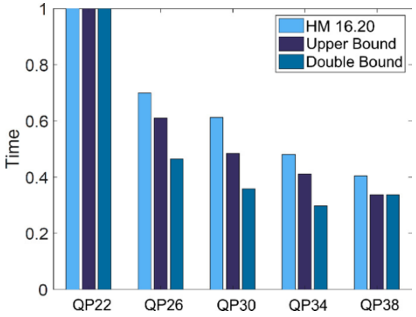
Fast Multi-rate Encoding for Adaptive HTTP Streaming, DCC 2020

UNIVERSITÄT KLAGENFURT

ATHENA Christian Doppler (CD) Laboratory

Video Coding for HAS

Fast Multi-rate Encoding

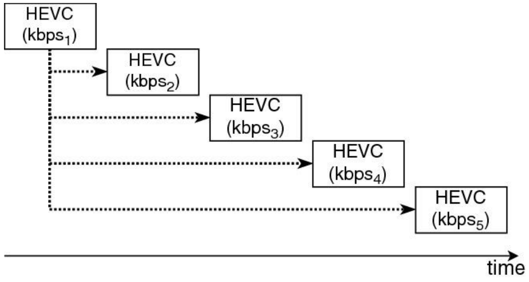


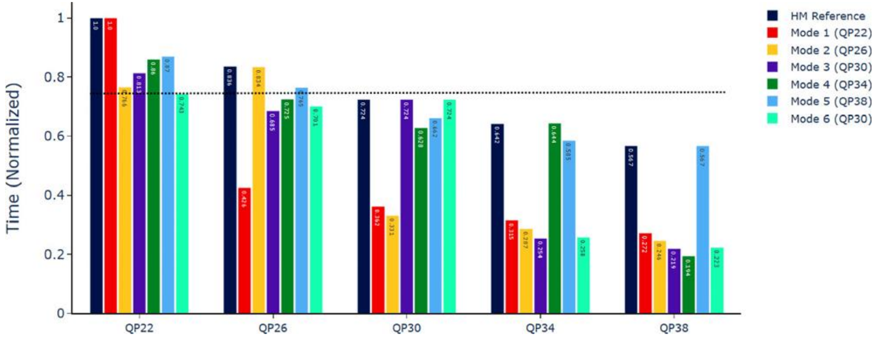
UNIVERSITÄT KLAGENFURT

ATHENA
Christian Doppler (CD) Laboratory

Video Coding for HAS

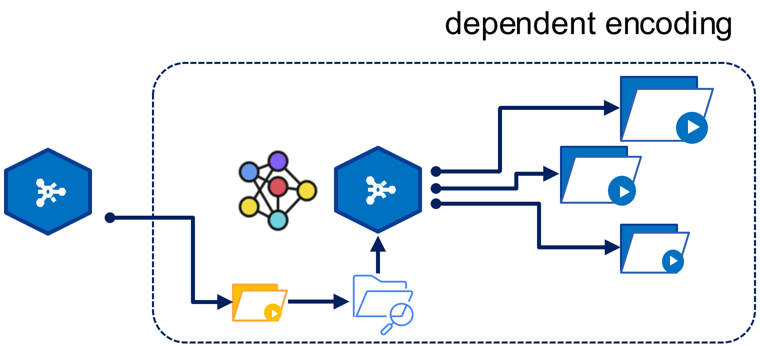
Fast Multi-rate Encoding





Video Coding for HAS

Fast Multi-rate Encoding



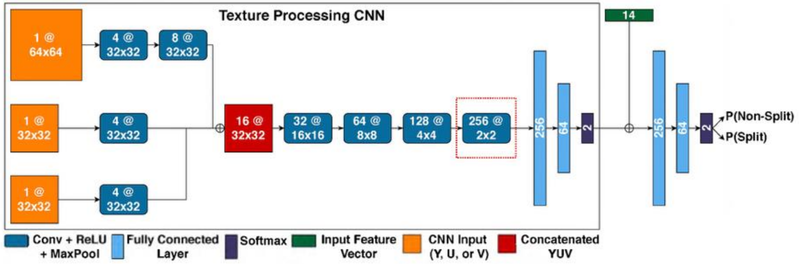


Fig. 2: Network architecture used for depth 0 classifier. The numbers in the boxes are in the following format from left to right: *Channel count* @ *Width*×*Height of channel* for convolution layers, *output size* for fully connected and softmax layers, and *input size* for the feature vector. Red dotted section is removed in the depth 1 classifier due to variance in the input size. Y, U, and V input sizes and intermediate channel sizes vary depending on the depth level (halved for depth 1 classifier).

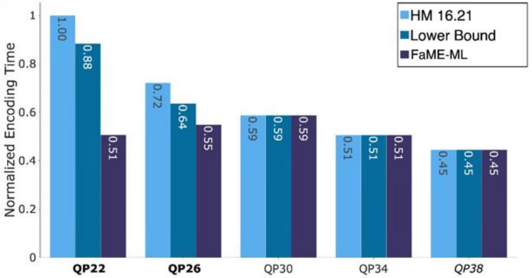
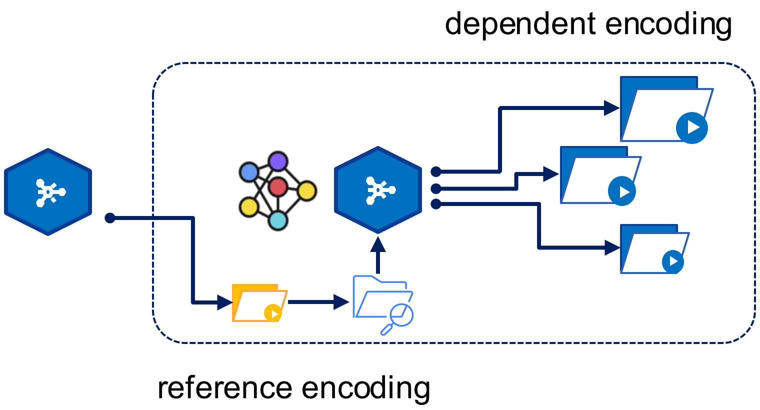


Fig. 3: Normalized average time-complexities in different QP levels using different methods.

Fast multi-resolution and multi-rate encoding for HTTP adaptive streaming using machine learning, OJSP 2021

Video Coding for HAS

Fast Multi-rate Encoding



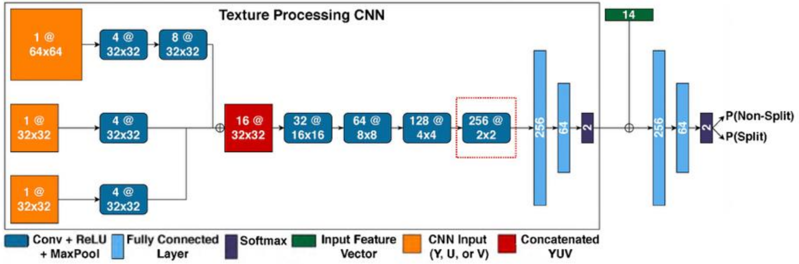


Fig. 2: Network architecture used for depth 0 classifier. The numbers in the boxes are in the following format from left to right: *Channel count* @ *Width*×*Height of channel* for convolution layers, *output size* for fully connected and softmax layers, and *input size* for the feature vector. Red dotted section is removed in the depth 1 classifier due to variance in the input size. Y, U, and V input sizes and intermediate channel sizes vary depending on the level (halved for depth 1 classifier).

TABLE I: Encoding Results for Test Sequences Using the Lower Bound and the FaME-ML

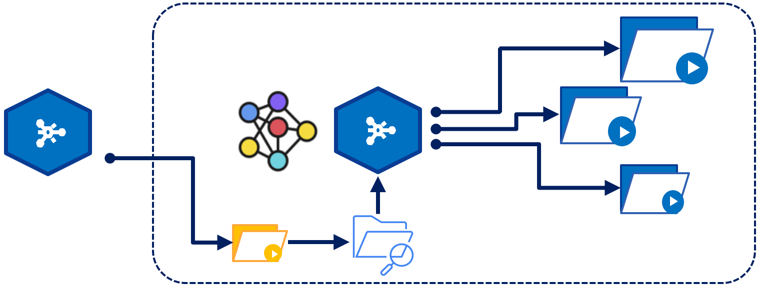
	Lower Bound					FaME-ML				
Sequence	ΔT	**BDR**$_P$	**BDR**$_P$ / ΔT	**BDR**$_V$	**BDR**$_V$ / ΔT	ΔT	**BDR**$_P$	**BDR**$_P$ / ΔT	**BDR**$_V$	**BDR**$_V$ / ΔT
DucksTakeOff	9.84 %	0.346 %	3.51	0.092 %	0.93	36.42 %	0.305 %	0.84	0.119 %	0.32
InToTree	3.11 %	0.368 %	11.83	0.688 %	22.12	54.59 %	1.325 %	2.42	0.511 %	0.93
OldTownCross	4.17 %	0.457 %	10.95	0.191 %	4.58	52.89 %	0.955 %	1.80	0.077 %	0.14
ParkJoy	21.23 %	0.404 %	1.90	0.083 %	0.39	36.04 %	0.920 %	2.55	0.250 %	0.69
RedKayak	12.72 %	0.764 %	6.01	0.282 %	2.21	22.98 %	0.525 %	2.28	0.184 %	0.81
RushFieldCuts	17.90 %	0.471 %	2.63	0.101 %	0.56	40.60 %	1.214 %	2.99	0.456 %	1.12
ControlledBurn	2.30 %	0.703 %	30.56	0.146 %	6.34	46.91 %	0.679 %	1.47	0.493 %	1.05
ParkRunning3	16.81 %	0.475 %	2.82	0.086 %	0.51	39.67 %	1.178 %	2.97	0.507 %	1.27
Average	**11.01 %**	**0.498 %**	**8.77**	**0.208 %**	**4.70**	**41.26 %**	**0.887 %**	**2.16**	**0.324 %**	**0.79**

Fast multi-resolution and multi-rate encoding for HTTP adaptive streaming using machine learning, OJSP 2021

Video Coding for HAS

Fast Multi-rate Encoding

dependent encoding



reference encoding

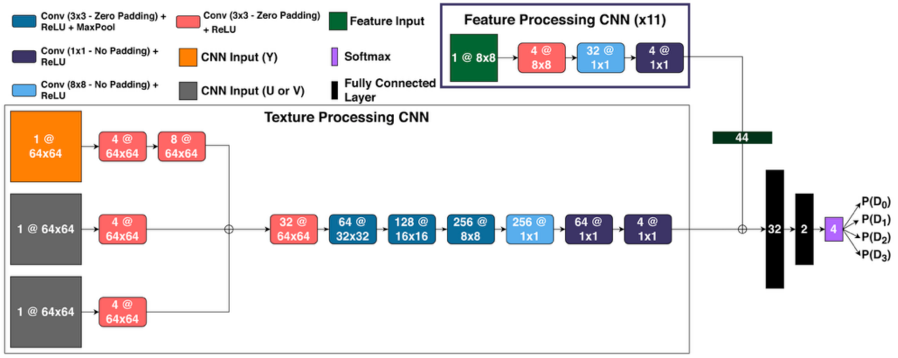


FIGURE 17. Modified FaRes-ML CNN architecture for minimum depth prediction. The intermediate output vectors now have 4 dimensions, one for each depth level.

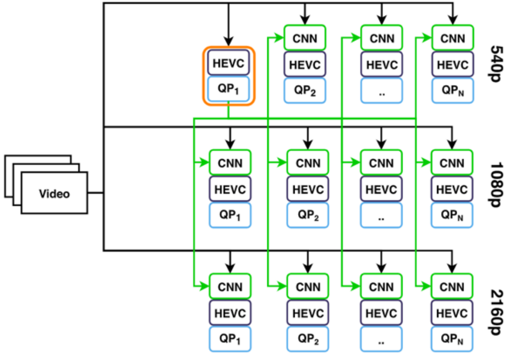


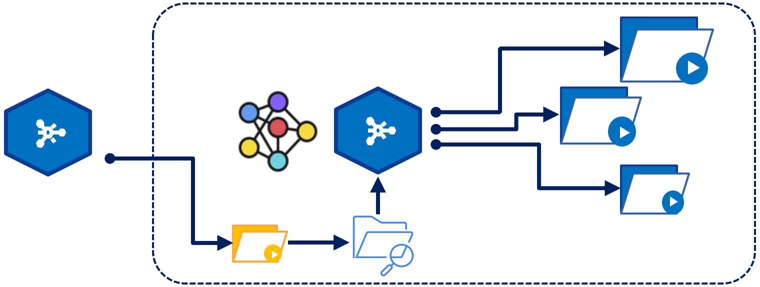
FIGURE 8. Flowchart of the FaRes-ML.

Fast multi-resolution and multi-rate encoding for HTTP adaptive streaming using machine learning, OJSP 2021

Video Coding for HAS

Fast Multi-rate Encoding

dependent encoding



reference encoding

Sequence		Lower Bound					FaRes-ML				
		ΔT	BDR_P	BDR_V	BD_{PSNR}	BD_{VMAF}	ΔT	BDR_P	BDR_V	BD_{PSNR}	BD_{VMAF}
3840x2160	Basketball	4.42 %	2.85 %	3.35 %	-0.084	-0.412	43.47 %	2.18 %	2.70 %	-0.068	-0.312
	Bunny	3.64 %	1.44 %	1.10 %	-0.051	-0.165	66.64 %	3.91 %	4.15 %	-0.139	-0.520
	Characters	2.62 %	6.92 %	4.84 %	-0.193	-0.095	53.76 %	2.76 %	1.44 %	-0.071	-0.079
	Contsruction	1.67 %	1.70 %	2.08 %	-0.047	-0.181	55.57 %	4.08 %	5.03 %	-0.095	-0.357
	Dolls	3.24 %	2.48 %	2.60 %	-0.061	-0.323	53.28 %	3.45 %	4.09 %	-0.081	-0.443
	Lake	2.74 %	0.14 %	0.04 %	-0.006	-0.011	42.46 %	1.76 %	1.57 %	-0.074	-0.274
Average		**3.06 %**	**2.59 %**	**2.34 %**	**-0.074**	**-0.198**	**52.53 %**	**3.02 %**	**3.16 %**	**-0.088**	**-0.331**
1920x1080	Basketball	5.12 %	0.98 %	0.90 %	-0.041	-0.175	45.27 %	3.47 %	3.14 %	-0.150	-0.538
	Bunny	3.15 %	0.50 %	0.49 %	-0.020	-0.031	60.52 %	2.57 %	2.96 %	-0.100	-0.324
	Characters	3.65 %	2.18 %	1.01 %	-0.076	-0.018	49.73 %	0.34 %	0 %	-0.011	-0.024
	Construction	2.12 %	0.94 %	1.03 %	-0.034	-0.156	55.23 %	2.09 %	2.41 %	-0.070	-0.192
	Dolls	4.42 %	1.10 %	1.57 %	-0.034	-0.236	49.67 %	3.93 %	4.73 %	-0.120	-0.580
	Lake	2.76 %	0.12 %	0.04 %	-0.004	-0.004	37.35 %	1.39 %	1.49 %	-0.054	-0.253
Average		**3.53 %**	**0.97 %**	**0.84 %**	**-0.035**	**-0.103**	**49.63 %**	**2.30 %**	**2.46 %**	**-0.084**	**-0.318**
960x540	Basketball	4.55 %	0.06 %	0.21 %	-0.003	-0.027	32.28 %	2.11 %	3.03 %	-0.102	-0.433
	Bunny	2.53 %	0.13 %	0.26 %	-0.006	-0.044	45.21 %	1.18 %	2.11 %	-0.052	-0.229
	Characters	2.84 %	0.31 %	0.75 %	-0.007	-0.013	39.44 %	0.01 %	0 %	-0.003	-0.038
	Construction	1.64 %	0.11 %	0.60 %	-0.004	-0.059	43.42 %	0.21 %	0.71 %	-0.007	-0.056
	Dolls	3.65 %	0.14 %	0.51 %	-0.005	-0.018	30.90 %	0.57 %	0.66 %	-0.020	-0.070
	Lake	2.50 %	0.02 %	0.29 %	-0.001	-0.036	28.66 %	0.89 %	2.35 %	-0.034	-0.317
Average		**2.95 %**	**0.12 %**	**0.43 %**	**-0.004**	**-0.032**	**36.65 %**	**0.83 %**	**1.48 %**	**-0.036**	**-0.191**
Total Average		**3.18 %**	**1.22 %**	**1.20 %**	**-0.037**	**-0.111**	**46.27 %**	**2.05 %**	**2.36 %**	**-0.069**	**-0.280**

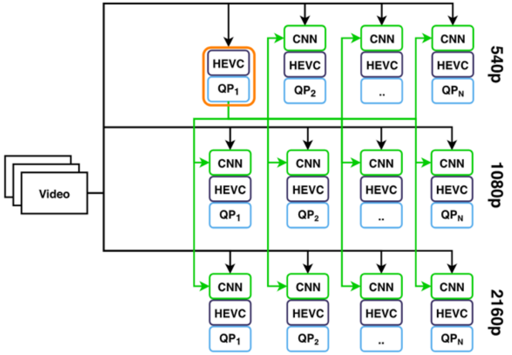
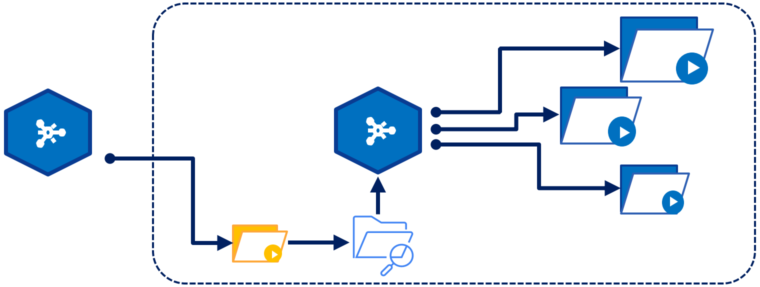


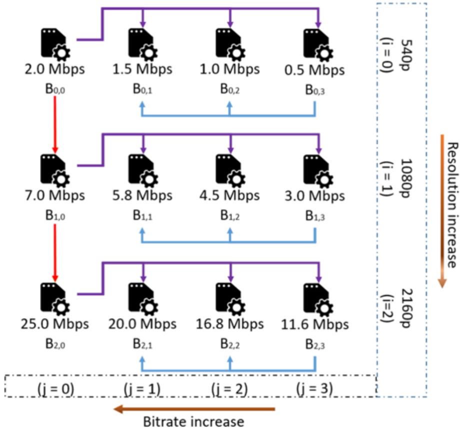
FIGURE 8. **Flowchart of the FaRes-ML.**

Fast multi-resolution and multi-rate encoding for HTTP adaptive streaming using machine learning, OJSP 2021

Video Coding for HAS

Fast Multi-rate Encoding





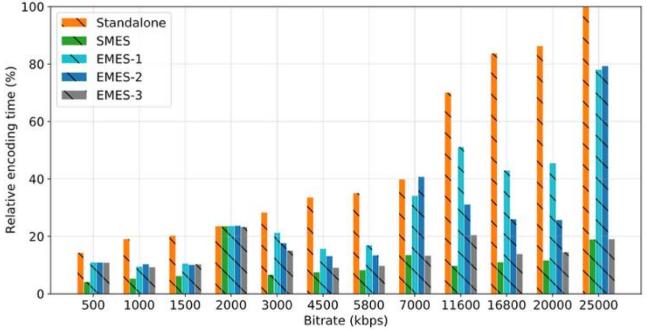


Fig. 12. Relative encoding time (in percentage) of all bitrate representations for the considered multi-encoding schemes. The encoding times are normalized by the stand-alone encoding time of the 25 Mbps representation.

EMES: Efficient Multi-encoding Schemes for HEVC-based Adaptive Bitrate Streaming, ACM TOMM 2023