Low Latency and Low Complexity: Industry Requirements for the Future Video Coding Standard H.267

Sergey Ikonin

VideoTech, 2025

HUAWEI

Speaker Introduction

Sergey Ikonin
Multimedia Communication Engineer

Research interests
- Video compression, Speech Compression
- Voice and Video over IP
- AI-based compression
- Joint source-channel coding

International standardization
- H.266/VVC, MPEG-5/EVC, JPEG AI
- Technical Expert and Contributor, Co-chair of Core Experiments and Ad Hoc Groups
- Ultra-low latency and packet loss resilience JVET AhG co-chair

Carrier
- Media Coding Lab Director at Huawei Technologies
- Head of Speech Lab, Spirit DSP

Outline

- Future video coding workshop on H.267
- Key industry requirements
- Low complexity
- Low latency
- AI-based compression
- Call for Evidence on capabilities beyond VVC

ITU/ISO joint workshop on "Future video coding"

Voice of industry

Workshop video record and materials are available:
https://www.itu.int/en/ITU-T/Workshops-and-Seminars/2025/0117/Pages/programme.aspx

Geneva, Switzerland, January 2025

Big workshop on Future Video Coding: 400 people on-site and more than 300 online

Session 1: Requirements and Use Cases: Voices from Industry and Users

Executives, managers and research leaders from the representative sectors of various industries gathered to give the audience a chance to hear their first-hand voices, which constituted the very basis for next generation video coding technical standards requirements. Industries chosen included smart phones and devices, Internet/OTT, social media, semiconductor, telecom, computing/cloud, automotive, etc.

Moderators: **Joern Ostermann**, Leibniz University Hannover and **Yuan Zhang**, Deputy Director, Big Data and AI Department, China Telecom Research Institute

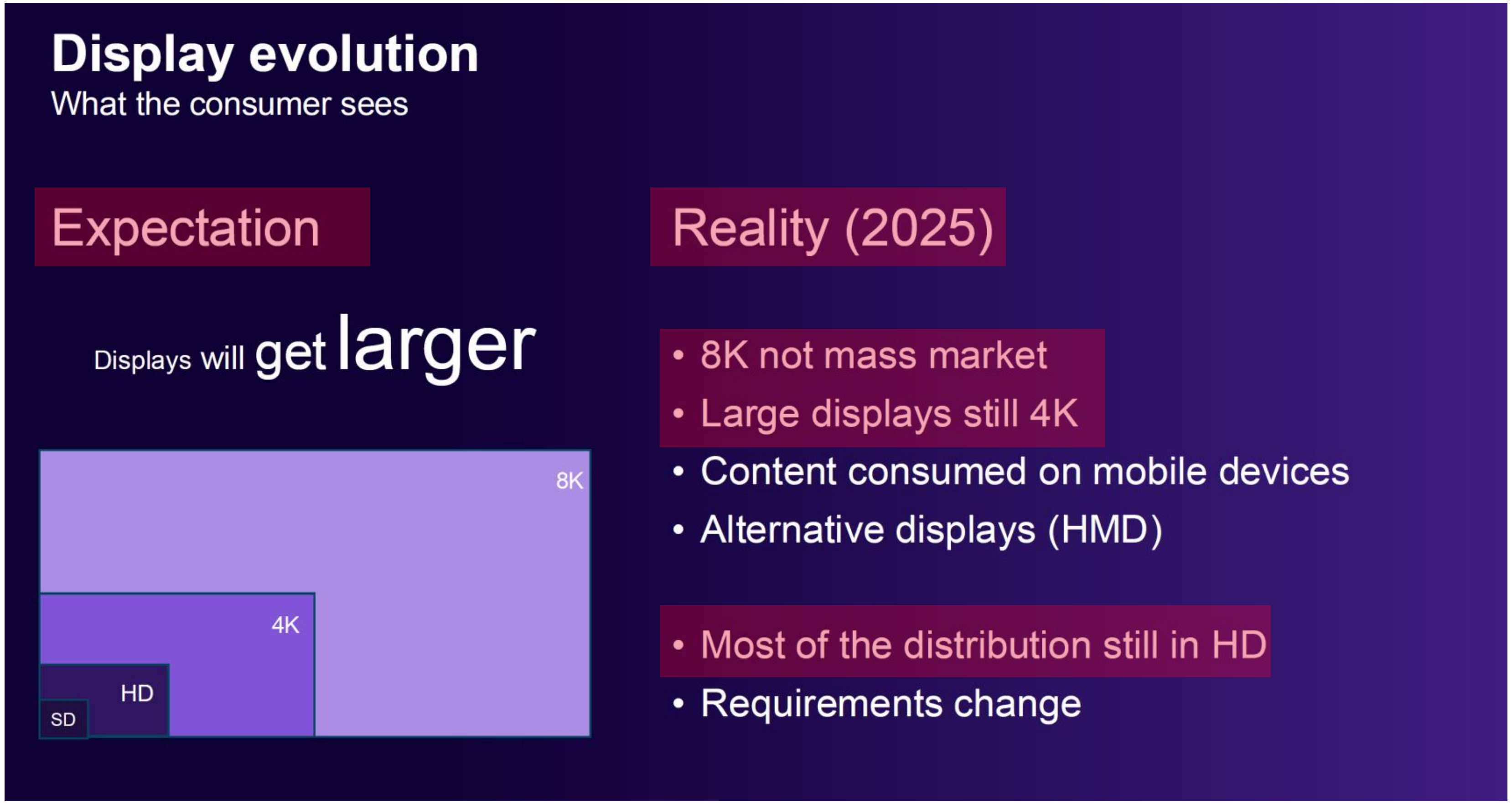
- **Kwang Pyo Choi**, Technical Vice President, Samsung Research, Samsung Electronics: *Industry Perspectives on Future Video Coding* [Presentation]
- **Andrew Segall**, Head of Video Coding Standards, Amazon: *Future Video Coding: An Amazon Perspective* [Presentation]
- **Li Lin**, Chief Expert, China Mobile: *Requirements for Video Encoding and Processing in the Multimedia Field* [Presentation]
- **Nicolai Otto**, Project Management Director, MainConcept GmbH: *Future Distribution Video Codec – Considerations from an Implementer* [Presentation]

Important conclusion:

- Compression rate is no longer the only goal for new standard
- Low latency and low complexity features are new features demanded

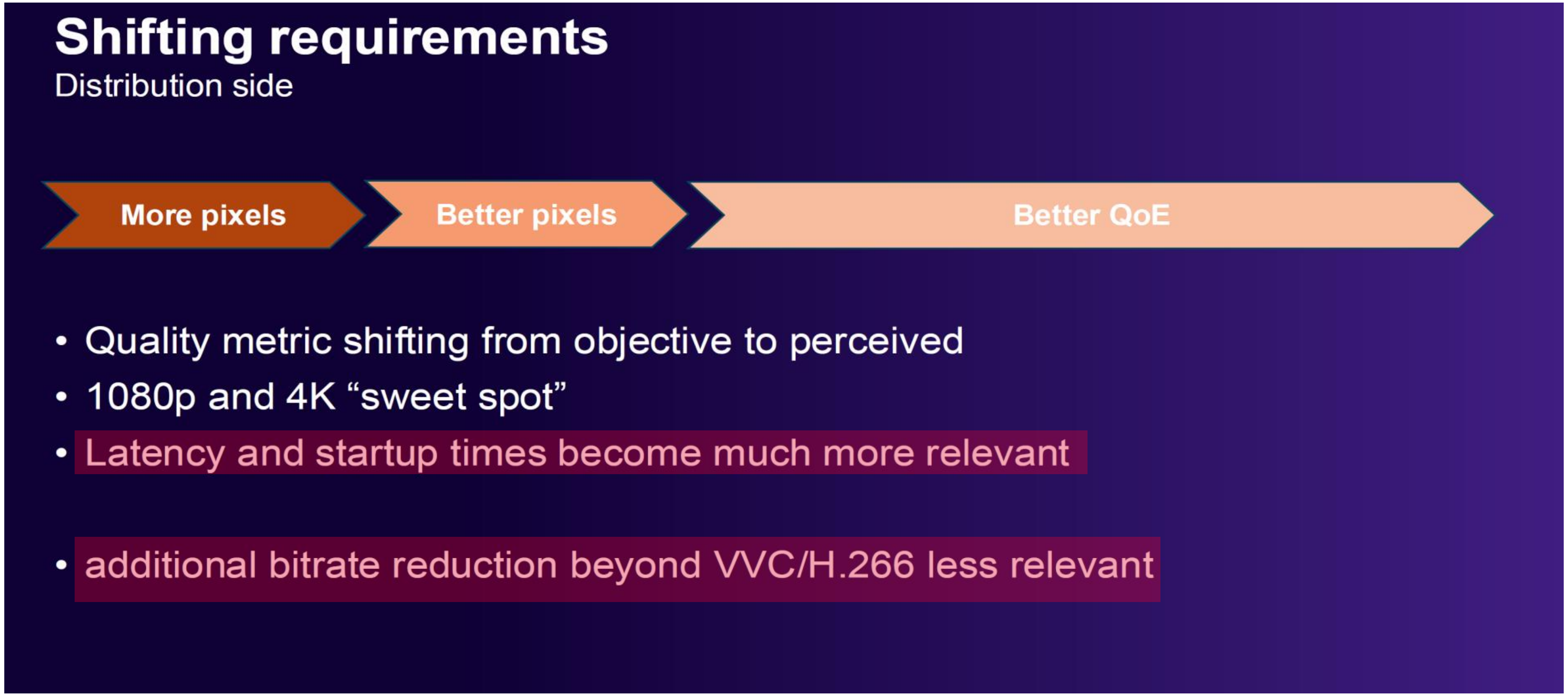
Voice of industry

Expectation and reality



Voice of industry

Shifting requirements



Voice of industry

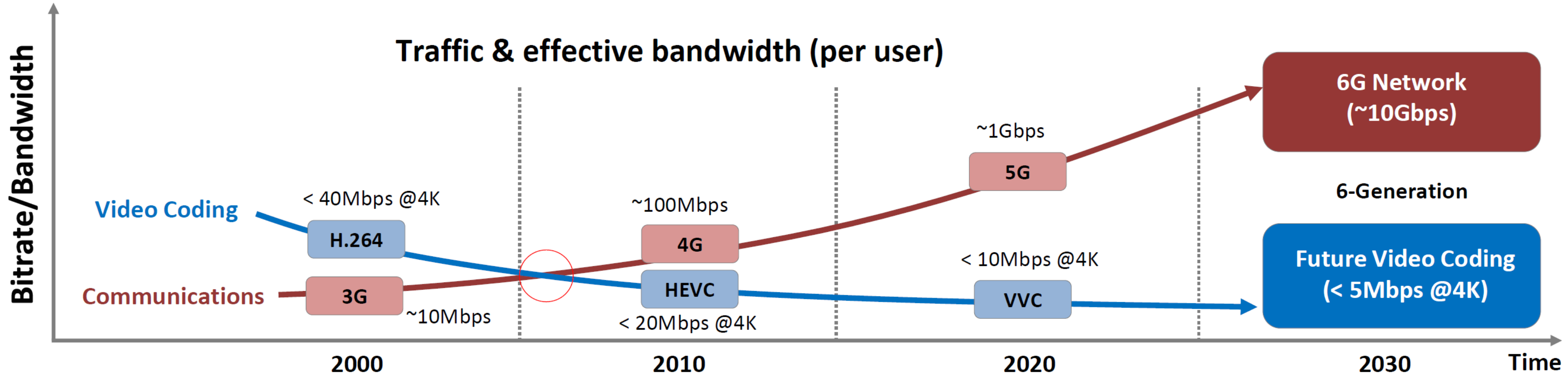
Needs for future video coding

Bitrate vs. Bandwidth

SAMSUNG

- **Bitrate reduction couldn't be a motivating factor for initiating a new codec standard**

	Video Coding Technology	Communication Technology
Evolution (over 10years)	Linearly increasing of coding gain (about 2x coding gain)	Dramatic increasing of bandwidth (about 10x bandwidth)
Technical limitation	Theoretical bound	Can use channel bonding
Deployment	Needs to change only end point	Need to change whole Infrastructure



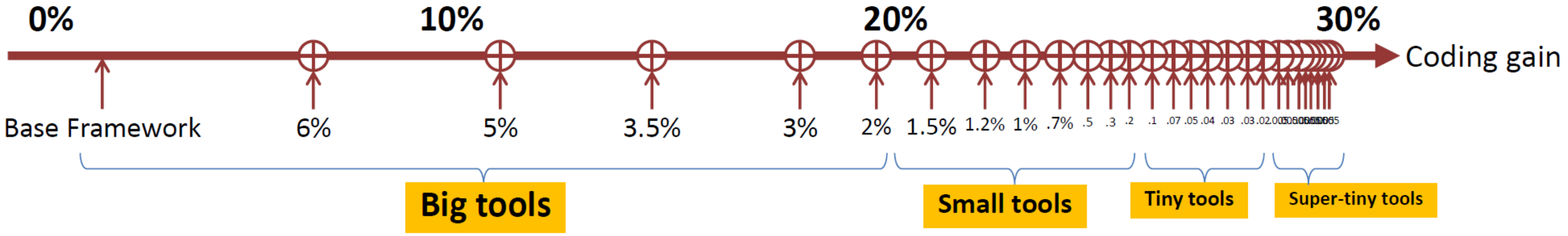
Voice of industry

2x times bitrate reduction cannot guarantee the market adoption anymore

SAMSUNG

No more stick to 30% coding gain

- **Again, hard to find market issue caused by lack of coding gain**
  - Standardizing a new video codec doesn't guarantee market adoption, only with 2x the coding gain, anymore
- **Improving coding gain seems to be harder than before**
  - Many experts say that video coding technology might reach theoretical limit
  - To improve small coding gain, huge computation is required for now
- **Avoid to adopt a tons of tiny coding tools to reach the 30% target goal**

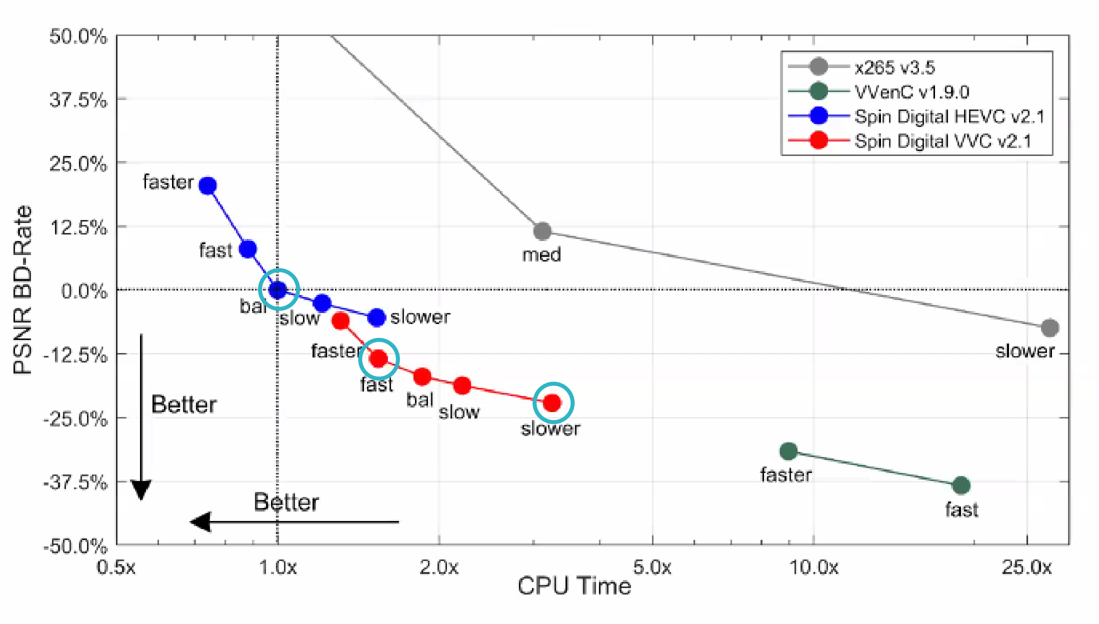


HUAWEI

Real-time encoder

Hard to get the same gain as on Test Model

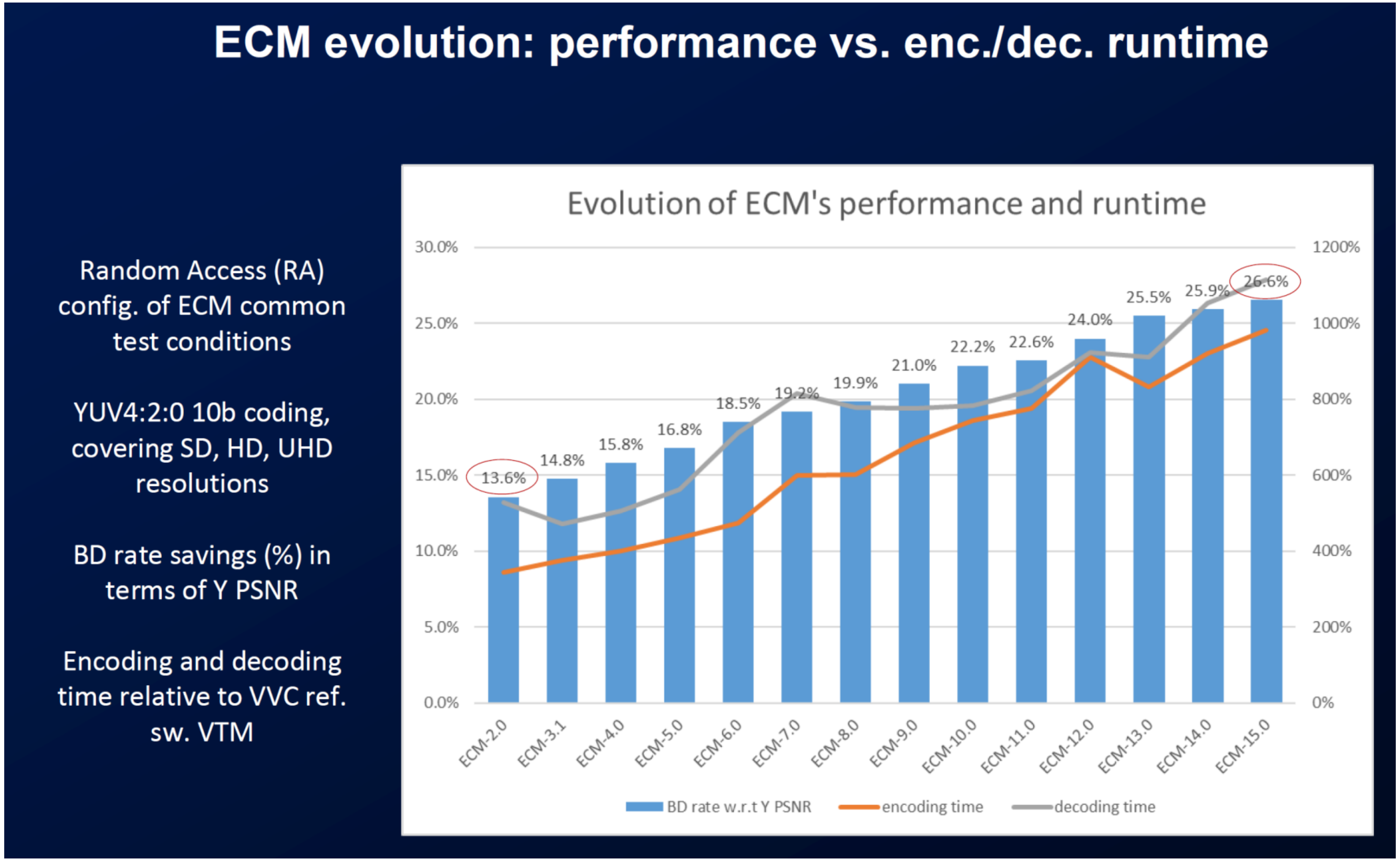
- It is more and more difficult to convert coding gain of test model to the coding gain in real-time configuration
- Real-time encoding H.265 vs H.264 → ~50%, H.266 vs H.265 → 10..15% , H.267 vs H.266 → ???



VVC over HEVC 'real-time' encoder from Spin Digital
- 13% compression gain with ×1.5 EncT. ↑
- 24% compression gain with × 3 EncT. ↑

JVET Enhanced Compression Model

Current performance of ECM beyond VVC

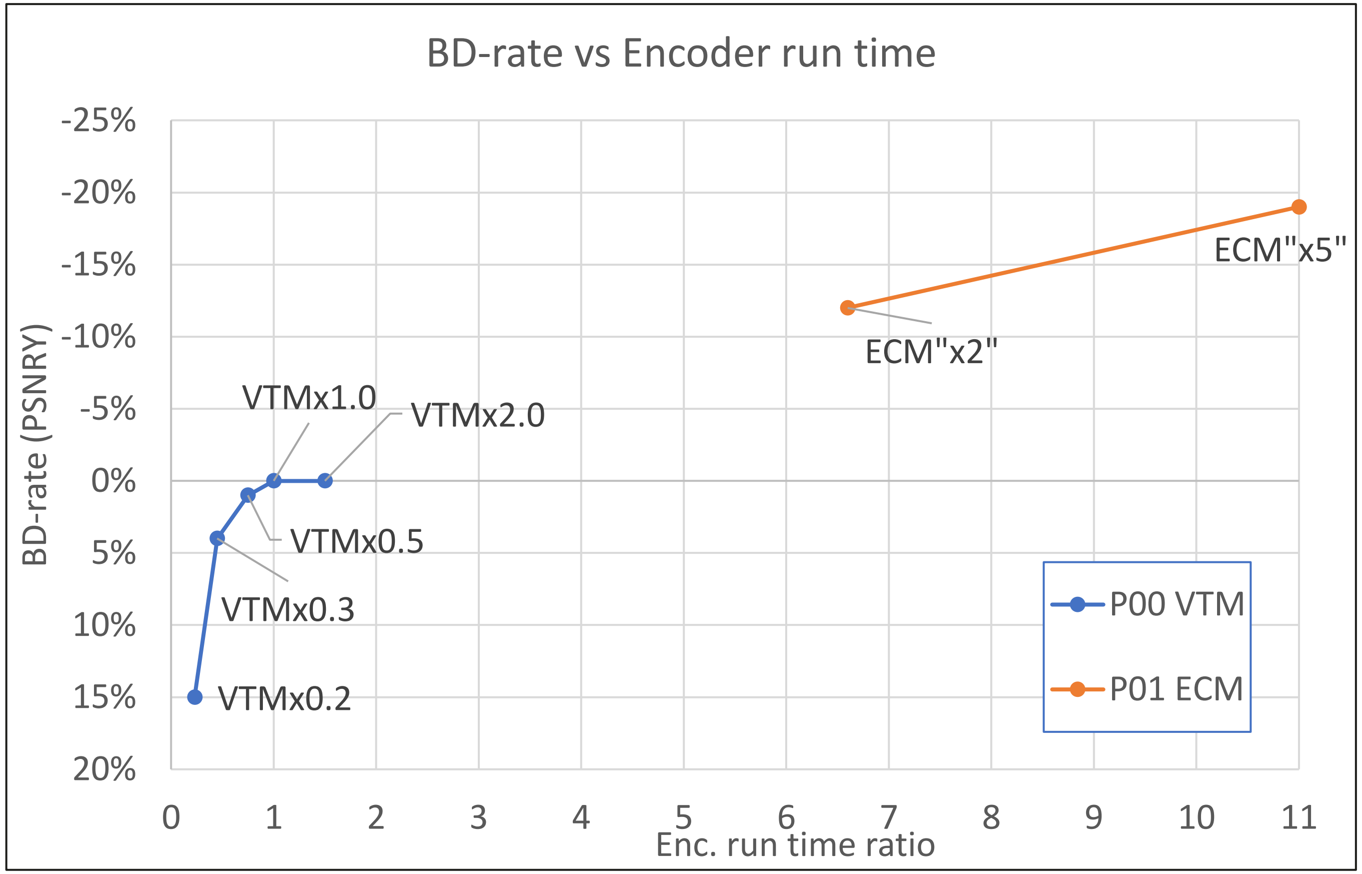


27% of coding gain at ~10x Encoder and Decoder complexity

JVET Call for Evidence

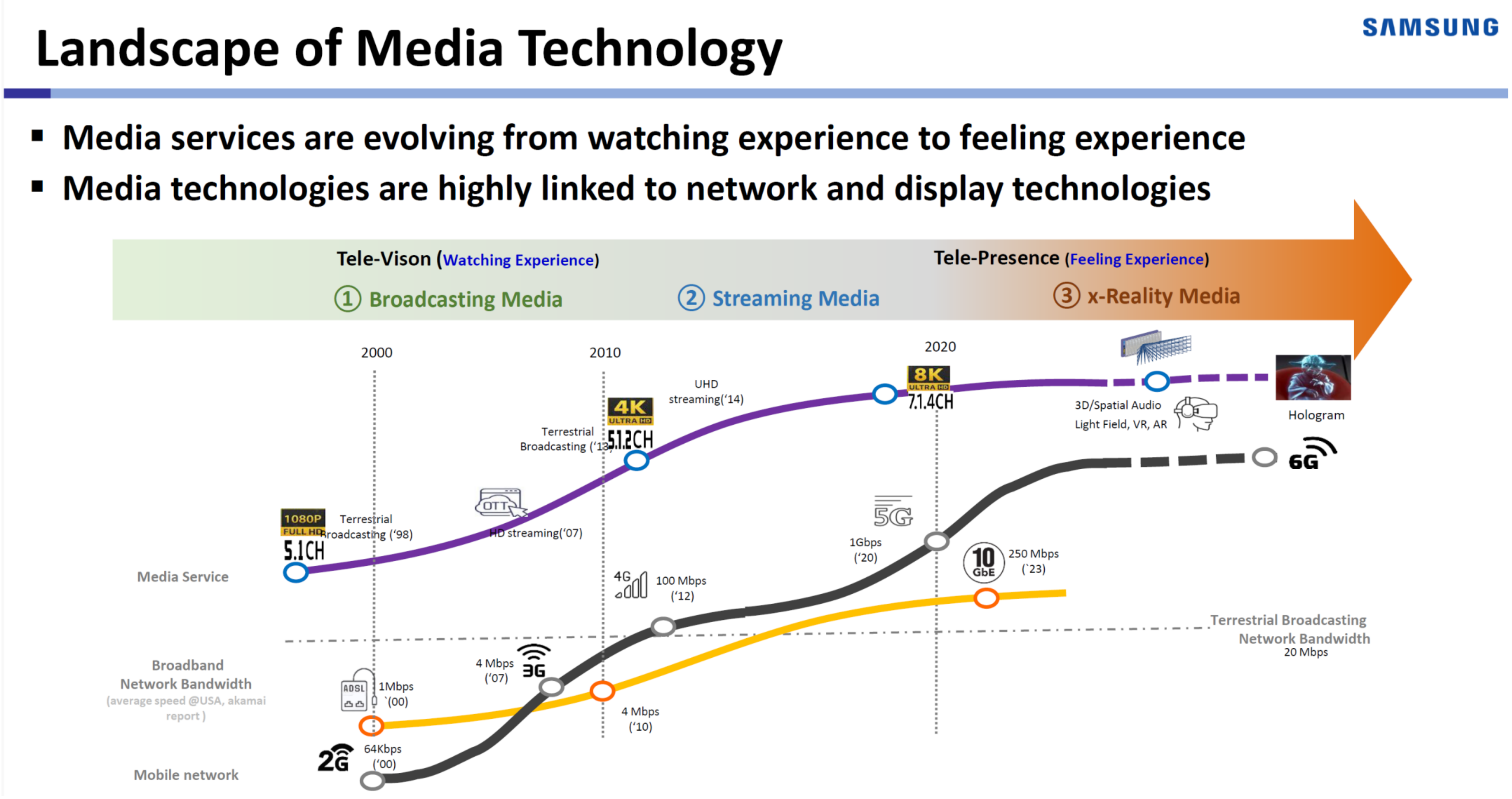
Constrained encoding runtime category

Results expected October, 2025, Geneva meeting
https://jvet-experts.org/



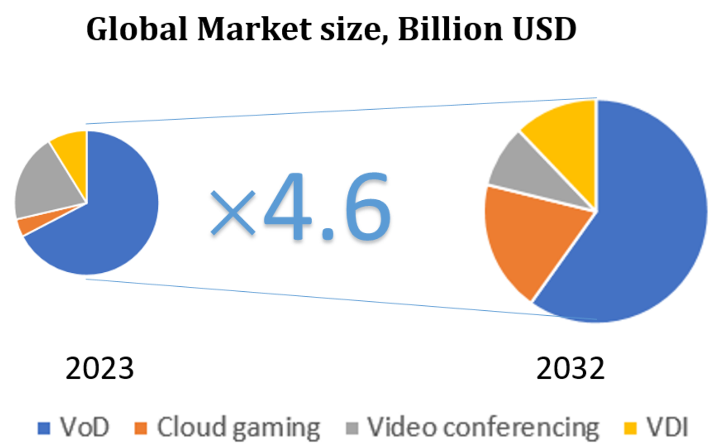
Voice of Industry

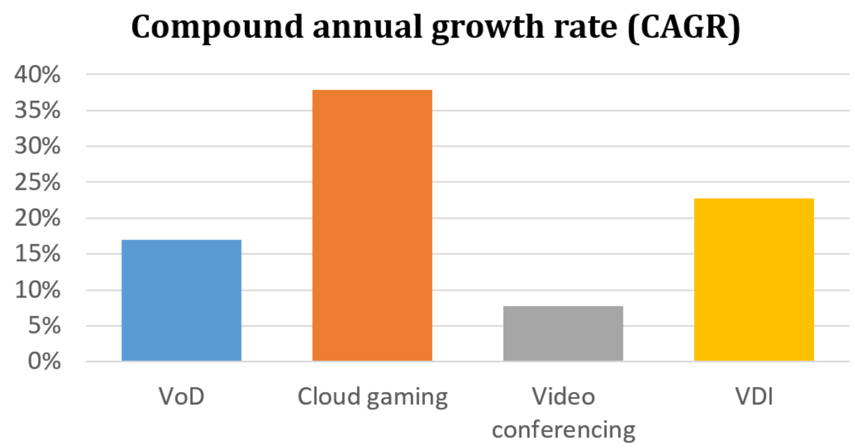
From “Watching” to “Feeling” experience



Interactive & low latency

Emerging highly-interactive application trends





CAGR is the average rate at which some value (investment) grows over a certain period of time

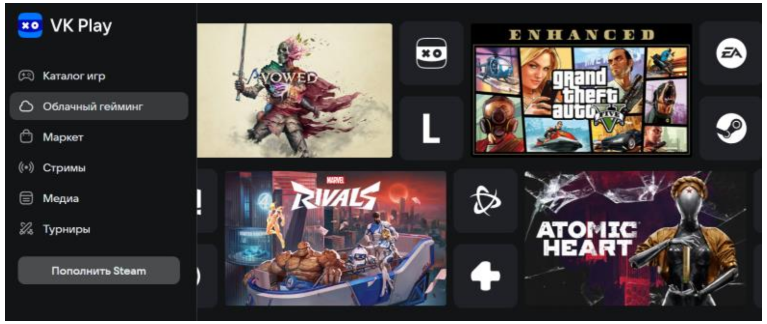
Source: https://www.fortunebusinessinsights.com

High resolution, high bitrates, and ultra-low latency

- **Cloud desktop:** 1080P~4K@60fps, 20~70Mbps, **50 ms**
- **Screen casting:** 4K@90fps, **10-20 ms**
- **Cloud gaming:** 2K@90fps, 50Mbps, **< 50 ms**
- **Cloud XR:** 8K@90fps, 100Mbps, **< 50 ms**

- Existing traditional applications could work somehow
- Latency minimization while keeping high bitrates and quality is still a challenge

Cloud gaming

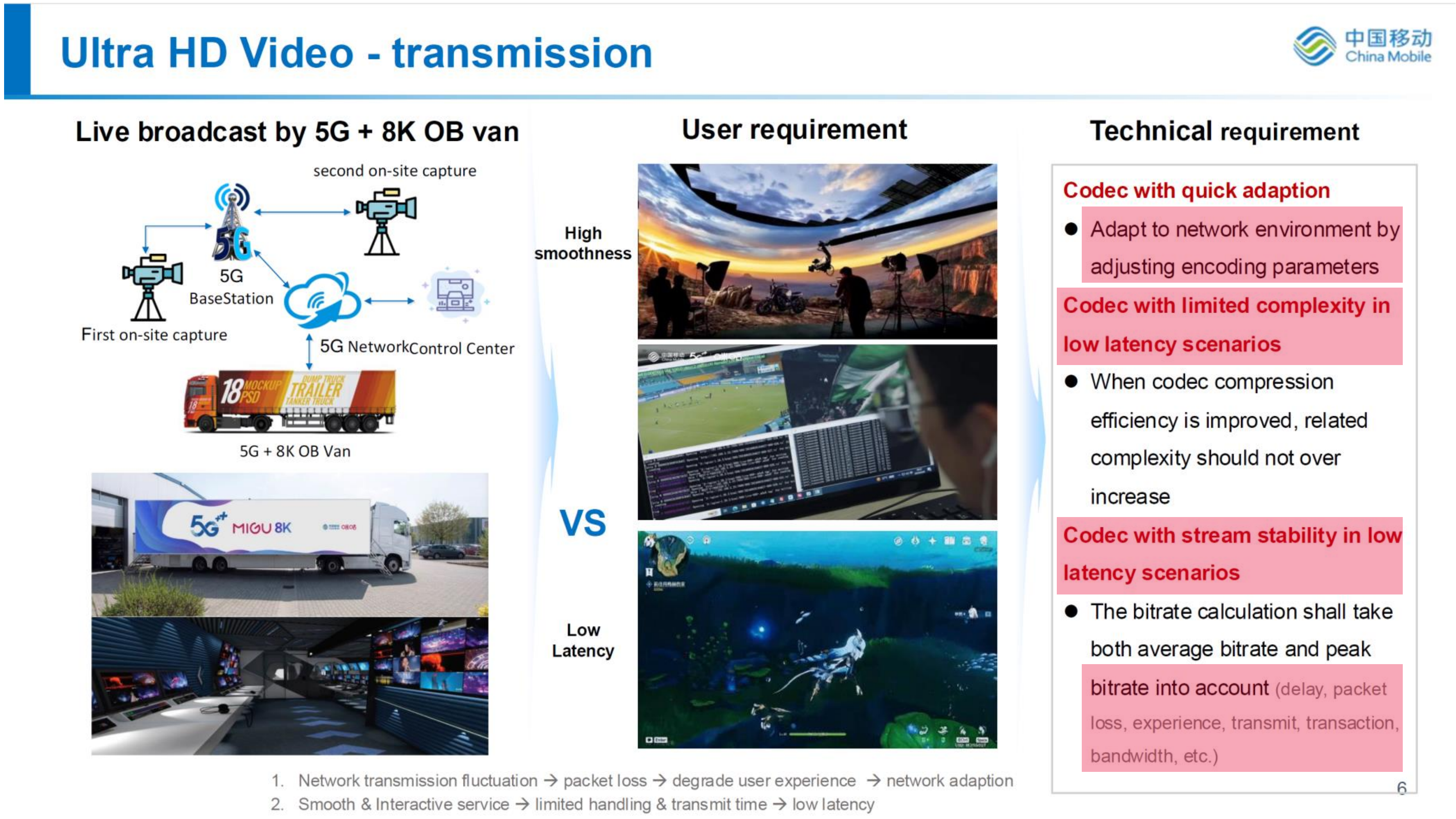


Stereoscopic 3D with new XR devices

Virtual Desktop Infrastructure

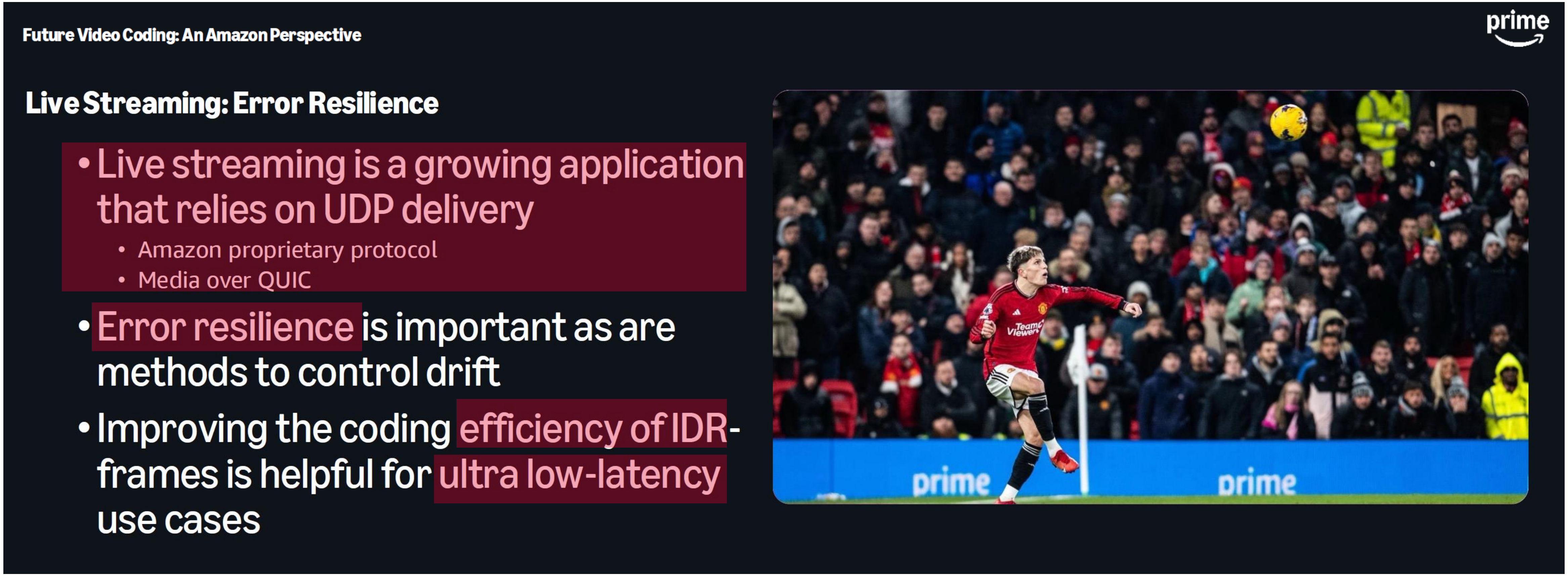
Voice of Industry

Network adaptation and stream stability in low latency scenarios



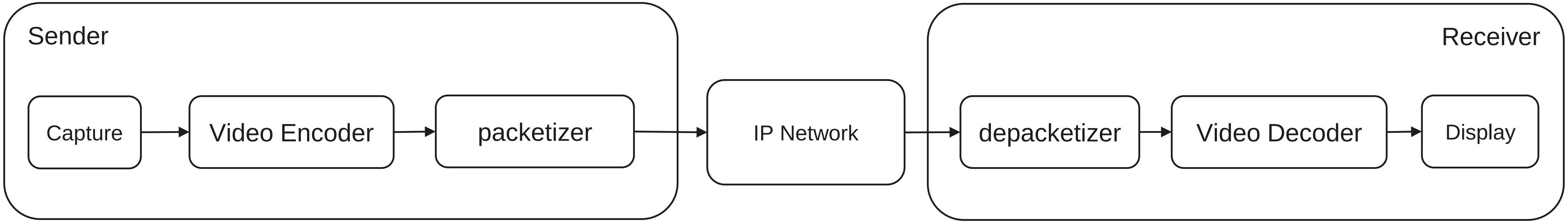
Voice of Industry

Error resilience decoding



Real-time video communication

Common architecture

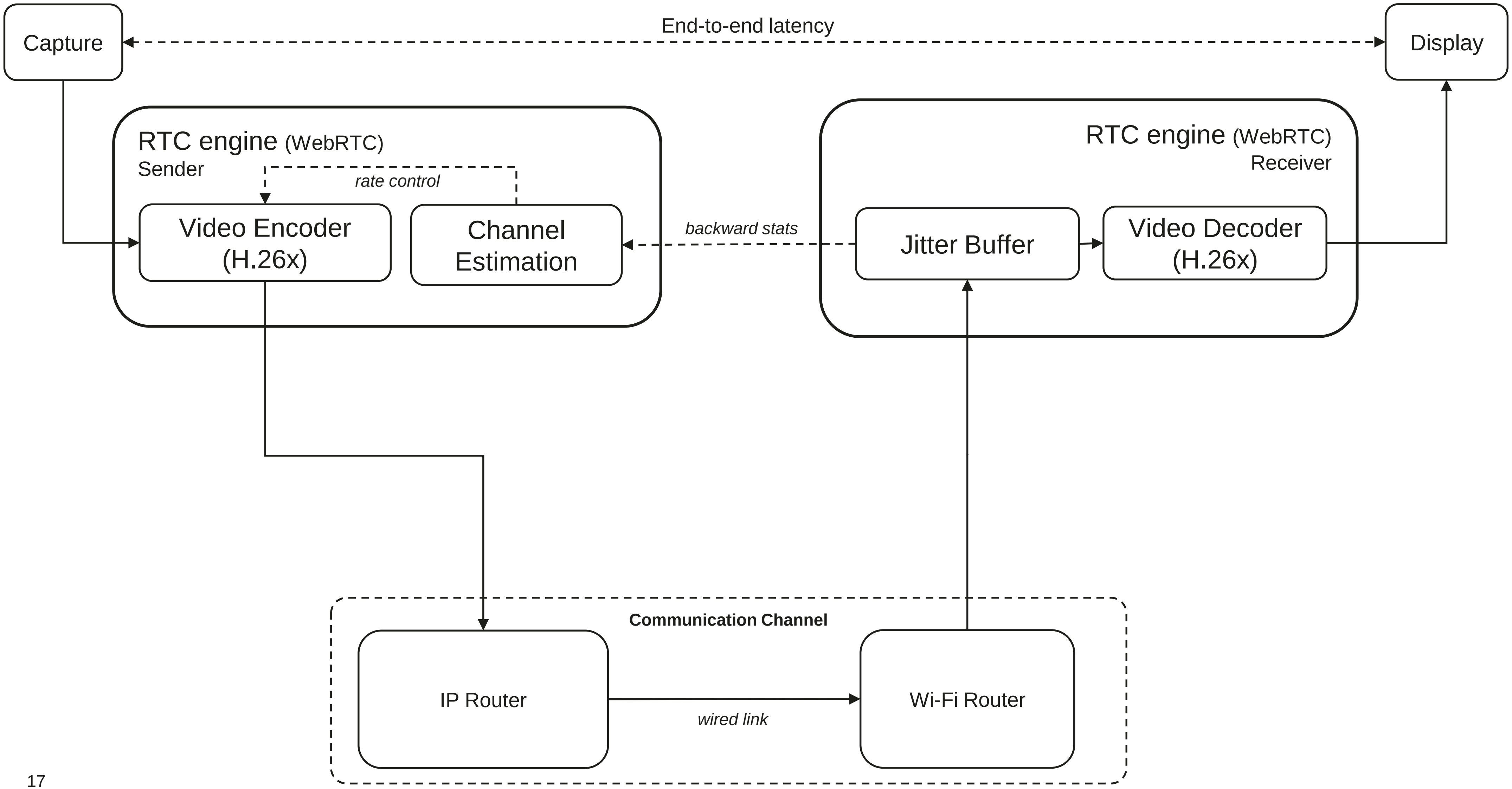


Network issues ⇒ **Video stalling**

- Some people say:
  "Our networks are good, and will be even better. No needs to care about packet loss and other network problems"
- During COVID time people's RTC experience was far for being perfect
- Personally, I stopped using Wi-Fi and switched back to old-good cable connection for RTC calls

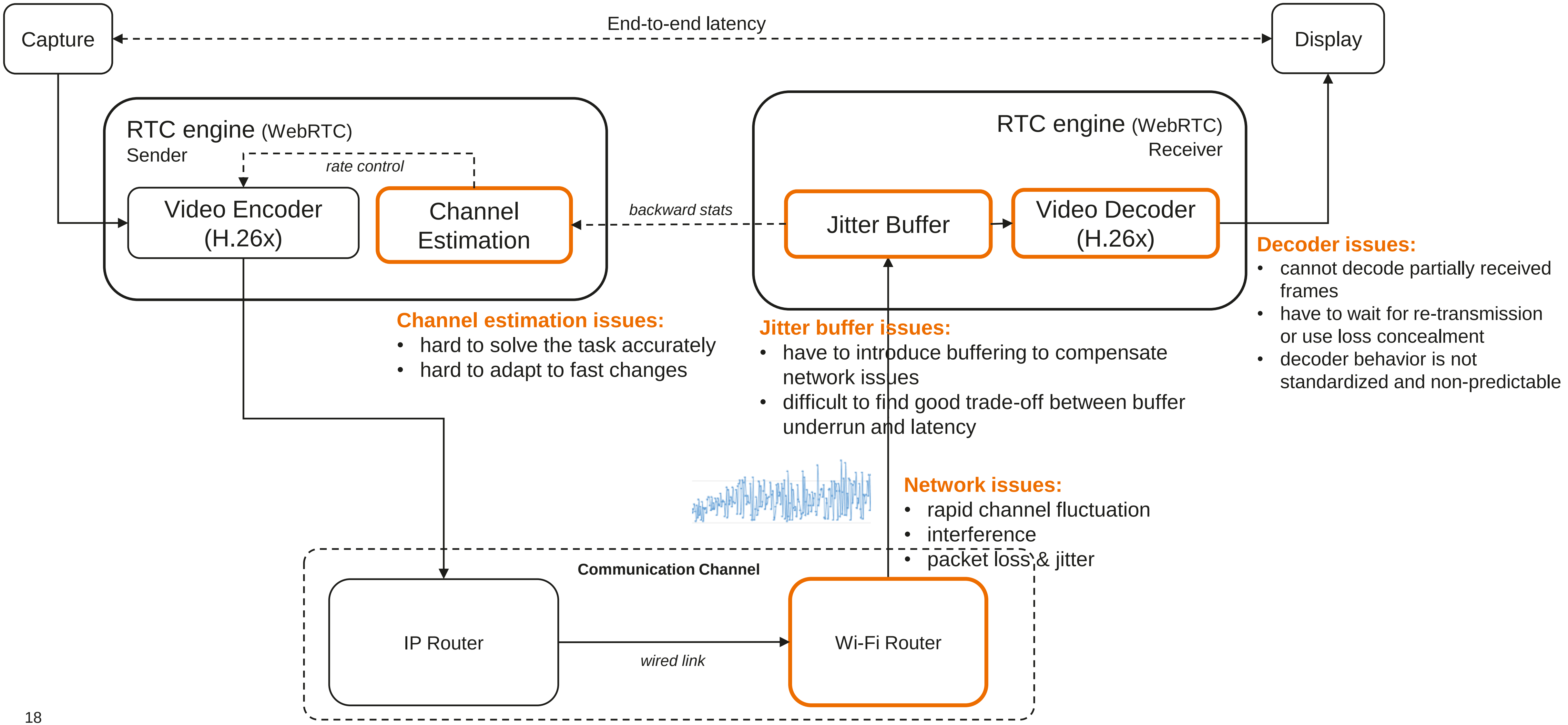
Real-time video communication

Common building blocks



Real-time video communication

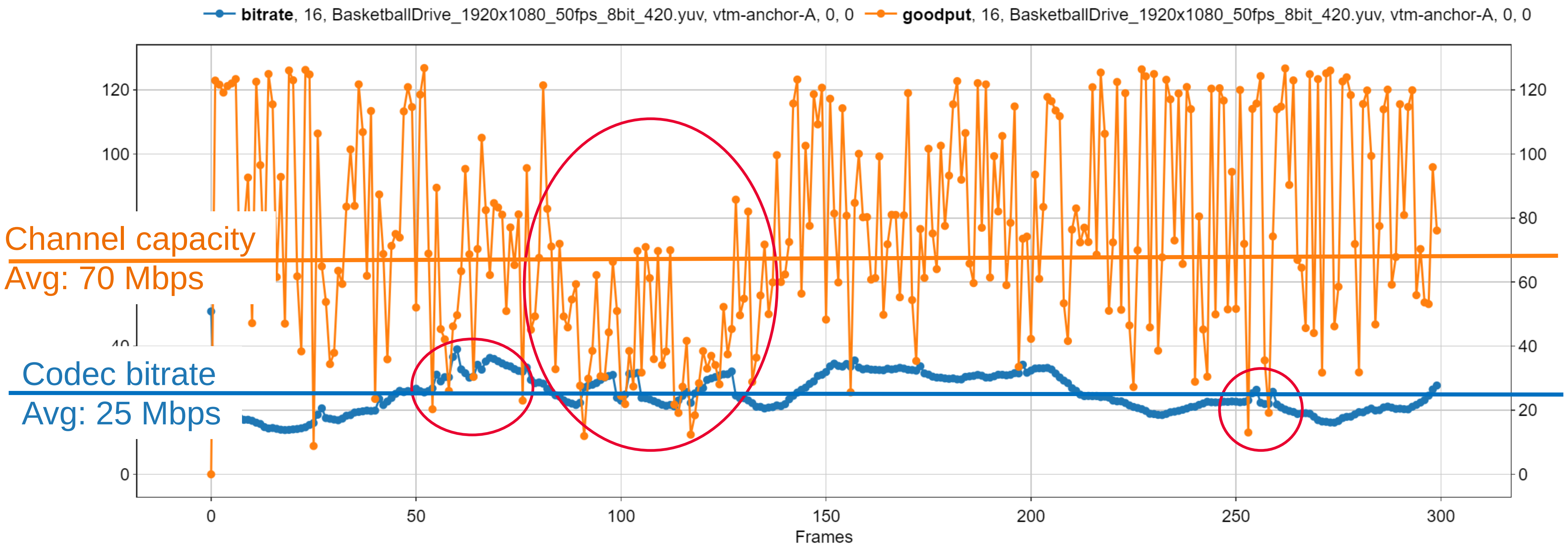
Common building blocks



Transmission problem analysis

Even x3 time higher channel bandwidth is not sufficient under ultra-low latency constrain

Channel capacity vs codec bitrate, example of Wi-Fi connection scenario



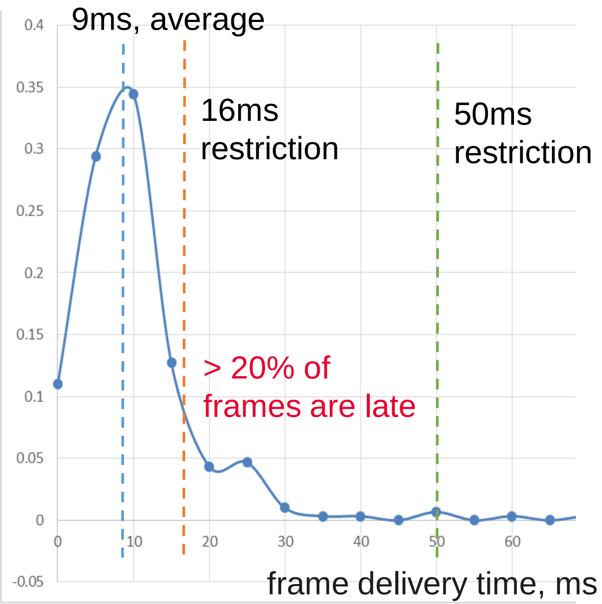
- Instant channel capacity is not sufficient for some frames
- At the same time, average bandwidth is ~3x times more than required

Transmission problem example

Video stalling depending on latency restriction

- BQTerrace, 1080p@60fps
- Frame interval: 16.6ms

Frame arrival time distribution:

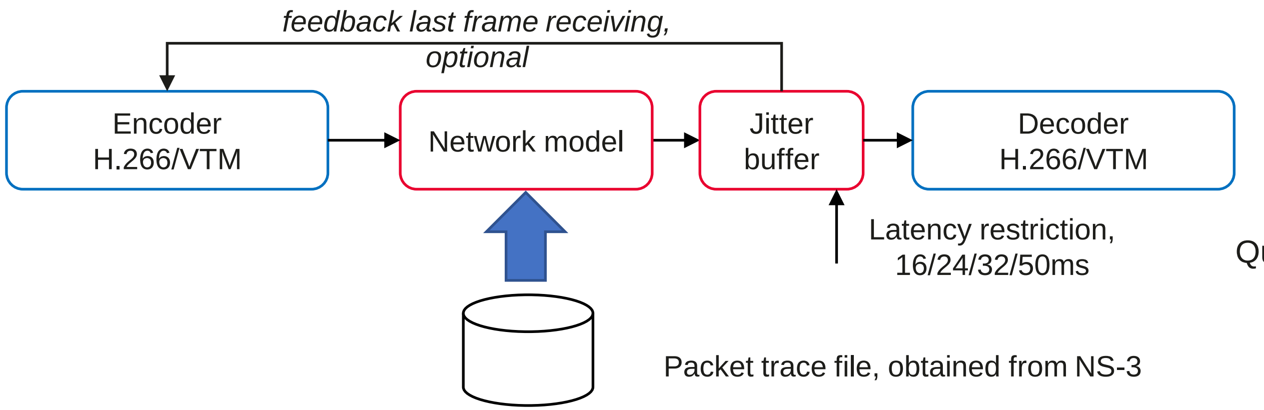


JVET AhG18 on Ultra-low latency

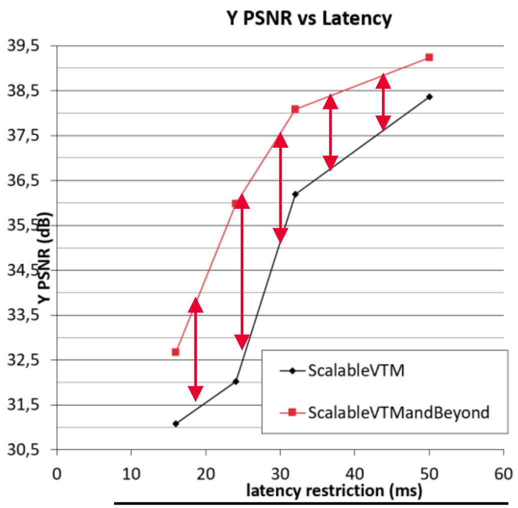
Established in Jan 2025

Software emulating low latency transmission problems for decoder

Available for MPEG members at: https://vcgit.hhi.fraunhofer.de/jvet-ahg-ull/VVCSoftware_VTM/



Quality metrics

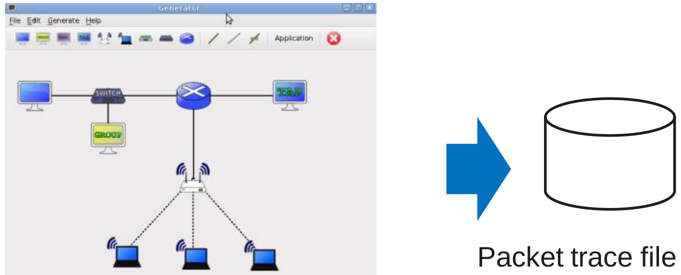


	Avg. PSNR		
	Y	U	V
Cactus seq	2,4 dB	1,4 dB	2,1dB

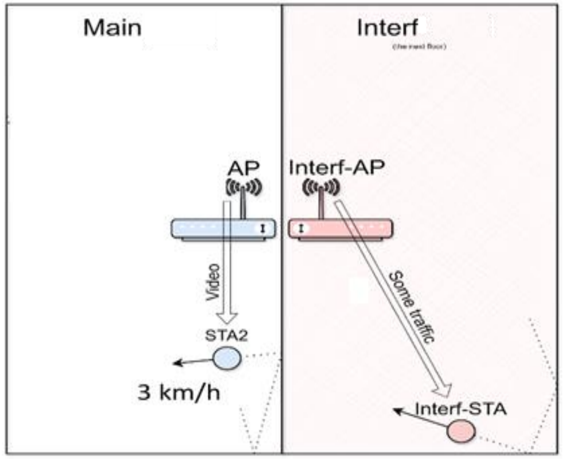
Quality metrics:
1. Average PSNR
2. Freeze ratio
3. Worst case PSNR, reconstruction quality of worst e.g. 5% or 10% or frames

- Packer trace file, simulating Wi-Fi connections is generated offline using NS-3 simulator, https://www.nsnam.org/

ns-3 Network Simulator



- Physical room typical for home/office environment



- 16m x 8m office room
- 1 Wi-Fi station (STA) connected to Access Point (AP) down streaming the video VVC encoded video.
- Wi-Fi parameters: 40 MHz bandwidth, TX power = 16 dBm (~40mW)
- Neighboring interfering Wi-Fi network is located in a room of same size.
- Interfering traffic is emulated by 50 Mbps neighborhood interference (~5 YouTube-flows of 1080p)
- STs are moving in their rooms with a speed of 3 km/h (walking speed).

ULL codec

Evidence of potential improvement on top of VVC – more than 2 dB PSNR improvement was demonstrated

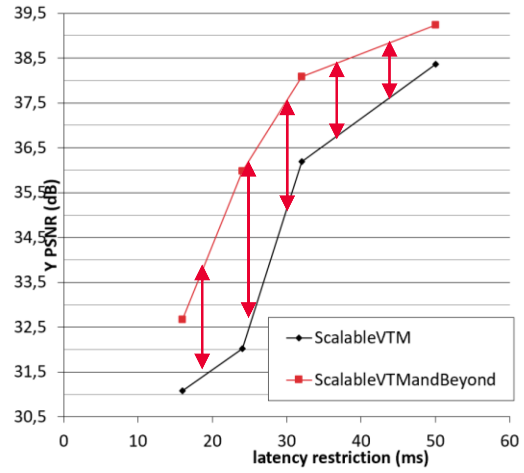
Average PSNR improvement at the same latency restriction

	IPPI w/o feedback, bypassIDR, bypass Base Layer			
	Over ScalableVTM, dB			
	Y	U	V	Freeze
Class A1, 4k	1,6	1,1	0,3	0,0%
Class A2, 4k	2,6	2,0	2,0	0,0%
Class B, 1080p	2,4	1,6	1,9	0,0%
Overall	**2,2 dB**	**1,6 dB**	**1,5 dB**	0,0%

Anchor: 2-layer scalable H.266

Test: Same as Anchor + Enhancements (see next slide)

Quality vs latency



	Avg. PSNR		
	Y	U	V
Cactus	2,4 dB	1,4 dB	2,1dB

- JVET contribution:

JVET-AM0218	AHG18: Solution beyond scalable coding for ultra-low latency and packet loss resilience	S. Ikonin, V. Khamidullin, I. Gribushin, M. Sychev, X. Ma, E. Alshina (Huawei)

Simulation conditions:

- Four latency restrictions experienced: 16, 24, 32 and 50ms
- Transmission priority model,
- Broadcasting scenario – no feedback from receiver

ULL codec

Subjective example and video demonstration

Demo videos available at:
https://shorturl.at/K1Woo

Demo videos, full link: https://drive.google.com/drive/folders/1a1q6JWOfg58hZg5-tAsdVDI_4FeFy3FA

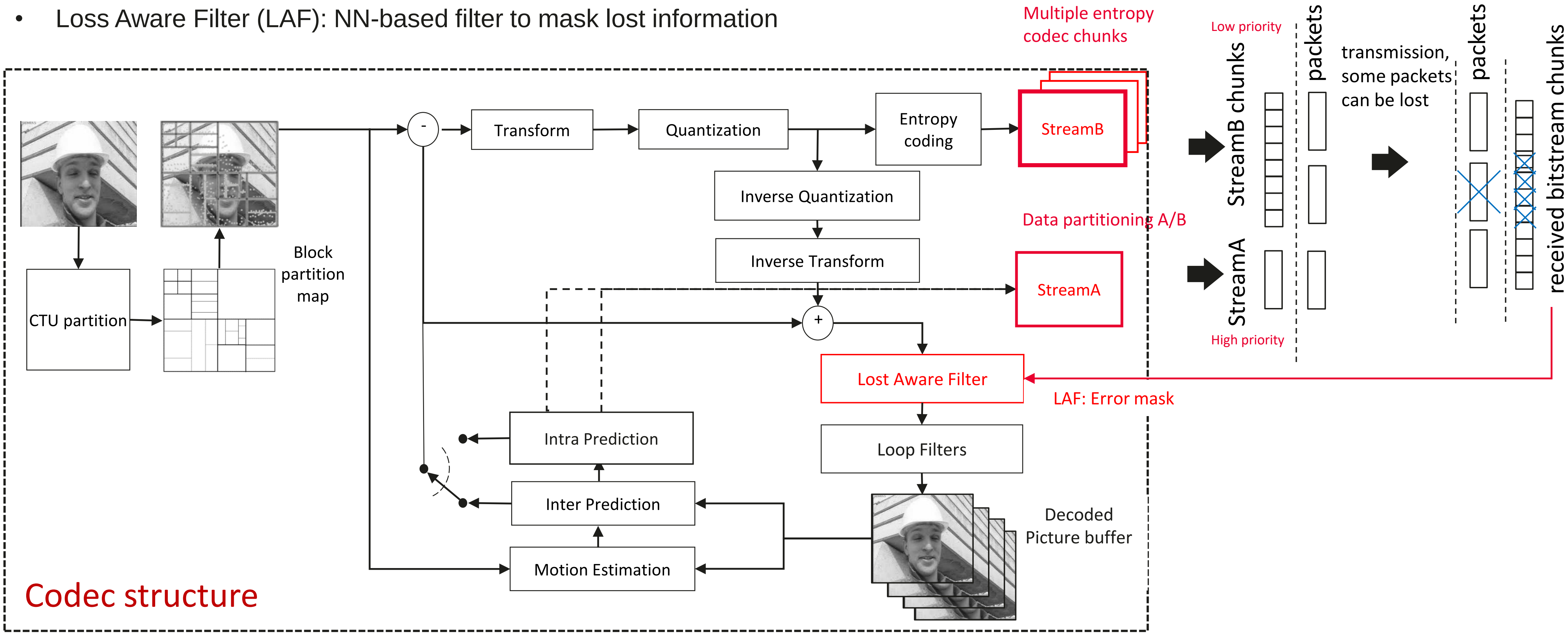
ULL codec

Key design elements

Source code is available: https://jvet-experts.org/

JVET-AN0045	AHG18: Codec software for ultra-low latency and packet loss resilience	S. Ikonin, V. Khamidullin, K. Malyshev, I. Gribushin, X. Ma, E. Alshina (Huawei)

- Data Partitioning – split prediction and residual information, reuse concept of H.264/3
- Multiple chunks of residual information, independently coded
- Loss Aware Filter (LAF): NN-based filter to mask lost information



ULL codec

Subjective examples of residuals loss effect, and masking by NN-based LAF filter

Lost residuals, 25.2 dB PSNR

LAF improvement, 27.8 dB PSNR

Lost residuals, 26.7 dB PSNR

LAF improvement, 36.8 dB PSNR

- Examples of reconstruction from partially received information
- If residual chunk is lost others can be used for reconstruction
- Lost residual blocks are further improved by LAF in-loop filter

Voice of industry

AI-based compression

Mega Trends of Video Coding Evolution

SAMSUNG

Media coding technology is evolving toward AI-based design for innovation coding gain

- Conventional signal processing-based compression algorithm is gradually reached out the coding gain limitation
- Media coding standard is adopting AI-powered coding tools, which show evidence to enable significant coding gain improvement

Conventional Video Codec

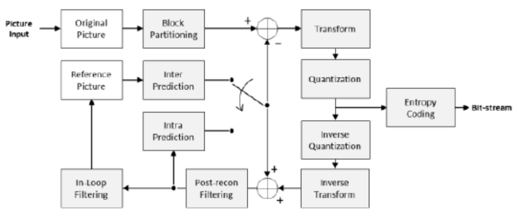
Enhancing coding tools through ML-based approach

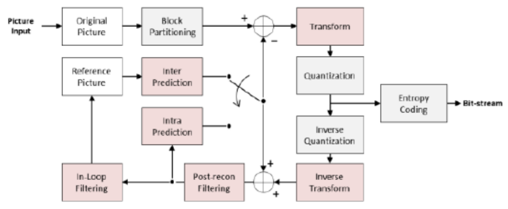
Design high coding gain tools and replace existing tool or add as a new tool

AI-based Video Codec

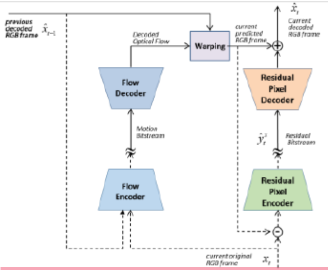
Novel architecture for next-generation compression technology

New design replacing the conventional coding algorithm used during 30 years





Trainable (ML-based) Non-trainable (conventional part)



Legacy Video Codec ('90 ~ '24)

Hybrid AI-Codec ('25 ~ '29)

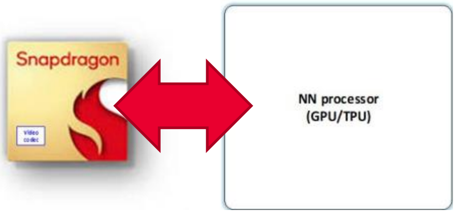
End-to-End AI-Codec ('30 ~)

H.265/HEVC ('13) H.266/VVC ('20) Future Video Coding ('30) E2E AI-based Future Video Coding ('35~)

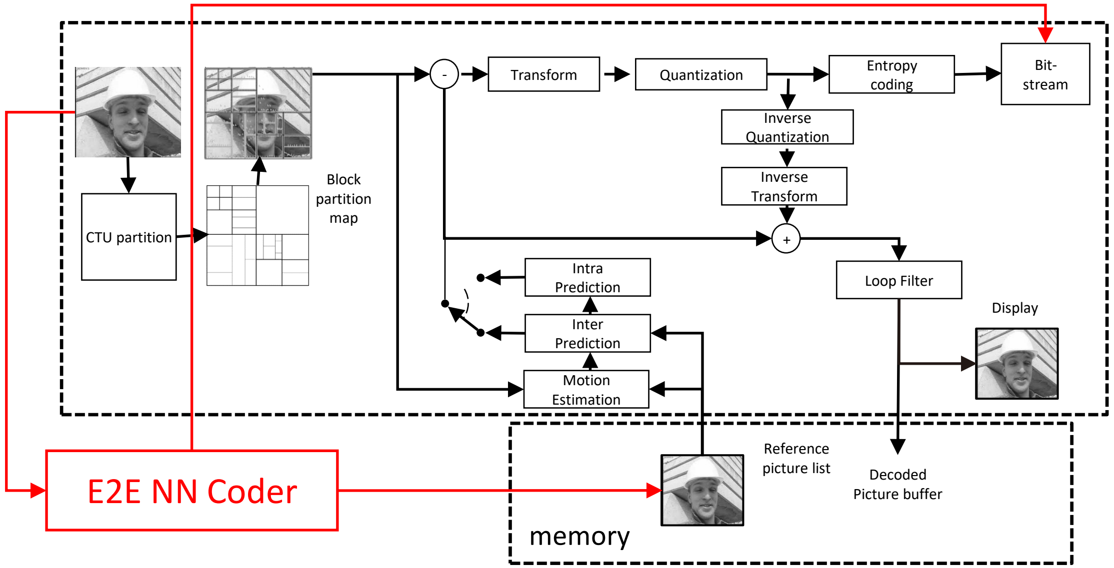
AI-based compression

Expectations and bottlenecks

- The biggest issue:
  data transfer to NN processor **on block-level**



- It is believed only tools with relatively
  high granularity of memory transfers can be acceptable



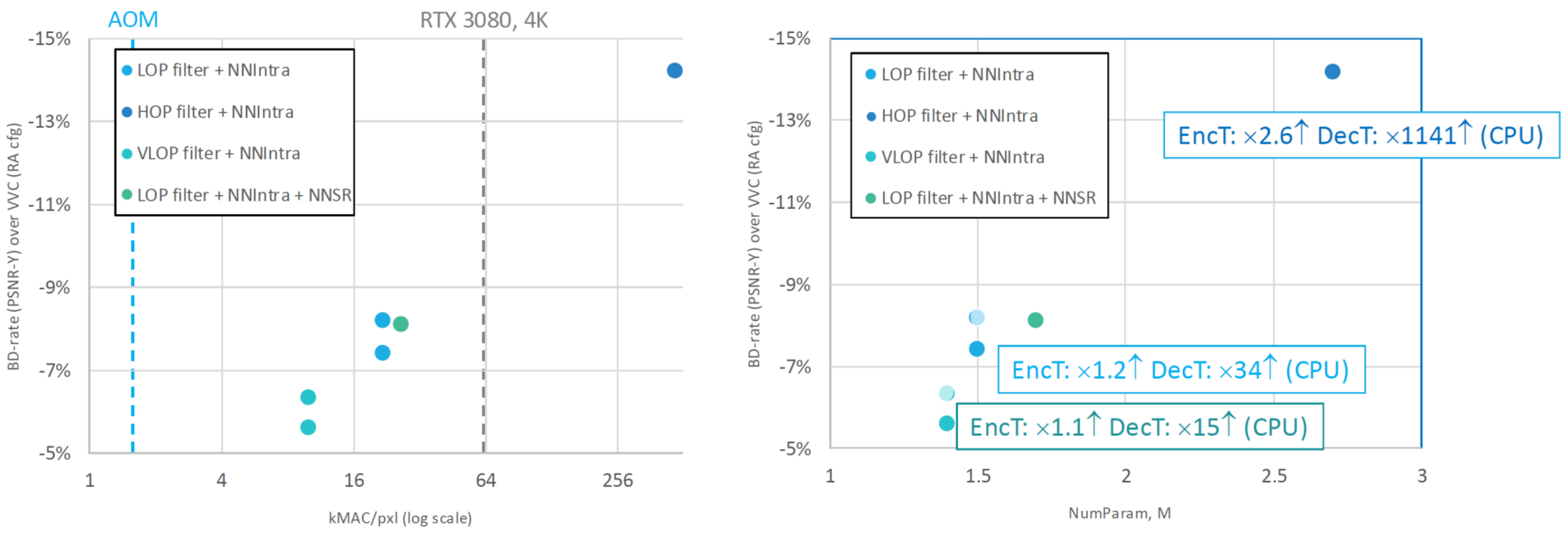
Tools/solutions with **frame-level memory-transfer**:

1. NN filter (post or in-loop)
2. Generated B-frame
3. Super-resolution
4. NN-based E2E coded intra/inter frame

JVET AHG11

Neural network-based coding

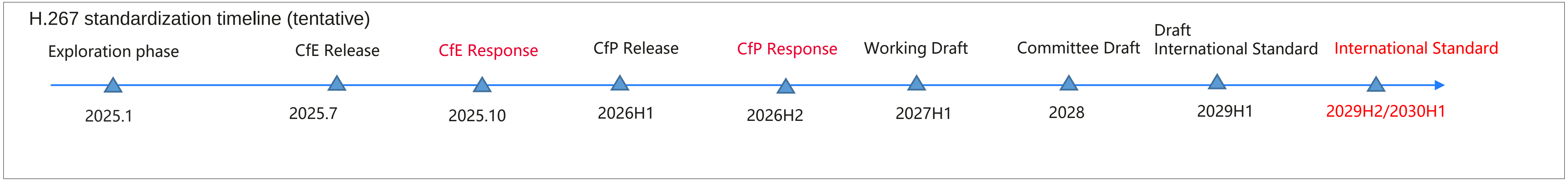
NNVC complexity performance trade-off



JVET Call for Evidence

CfE document JVET-AM2026
https://jvet-experts.org/

CfE response is the nearest mailstone expected in JVET October meeting



CfE categories:

- General video compression improvement
- Constrained encoding compression improvement
- Additional functionality as Ultra-low latency and packet loss resilience

Conclusion

- Compression efficiency not a sole priority for the new standard
- Low complexity and implementation cost is specific requirements
- Low latency communication is recognized as important for future applications
- Hybrid architecture enhanced by frame level AI-based tools is anticipated
- Call for Evidence responses will show the potential

Stay tuned: http://jvet-experts.org

Thank you.

Sergey Ikonin
Multimedia Communication Engineer

sergey.Ikonin@huawei.com
sergey.Ikonin@gmail.com