CREATE
CONNECT
LIVE
inspire

Comparison of Video Coding solutions

Michel Kerdranvat (InterDigital)

interdigital™

© 2020 InterDigital, Inc. All Rights Reserved.

Outline

- **Introduction**
- **Overview of algorithms of HEVC, VVC, EVC, AV1 and AVS3**
- **Test conditions for HEVC, VVC, EVC and AV1**
  - Coding solutions in test
  - Scenarios: Broadcast & Streaming
- **Results on objective metrics: PSNR, VMAF and MS-SSIM**
  - HD and 4K
  - 8K
- **Conclusion**

© 2020 InterDigital, Inc. All Rights Reserved.

Introduction

- **AVC/H.264** [2003] is still the dominant codec
- **HEVC/H.265** [2013] developed for UHDTV services
- **AV1** [2018] developed by Alliance for Open Media (AOM)
- **VVC/H.266** [2020] developed for immersive video services:
  - 8K, HDR, Virtual Reality with 360° video , mobile…
- **EVC** [2020] developed in MPEG only (not ITU-T)
- **AVS3** [2020] developed in China

© 2020 InterDigital, Inc. All Rights Reserved.

Introduction

- **Can a meaningful performance comparison of Video Coding solutions be made?**
  - Different reference softwares
  - Different configuration parameters
  - Different algorithms
  - Different picture types or GOP structures
  - Different test video sequences …
- **This presentation is based on a paper published in The ITU Journal: « The future of Video immersive media »**
  - https://www.itu.int/en/journal/2020/001/Pages/09.aspx
  - « The video codec landscape in 2020 »
    - Michel Kerdranvat, Ya Chen, Rémi Jullian, Franck Galpin, Edouard François InterDigital R&D, Rennes, France

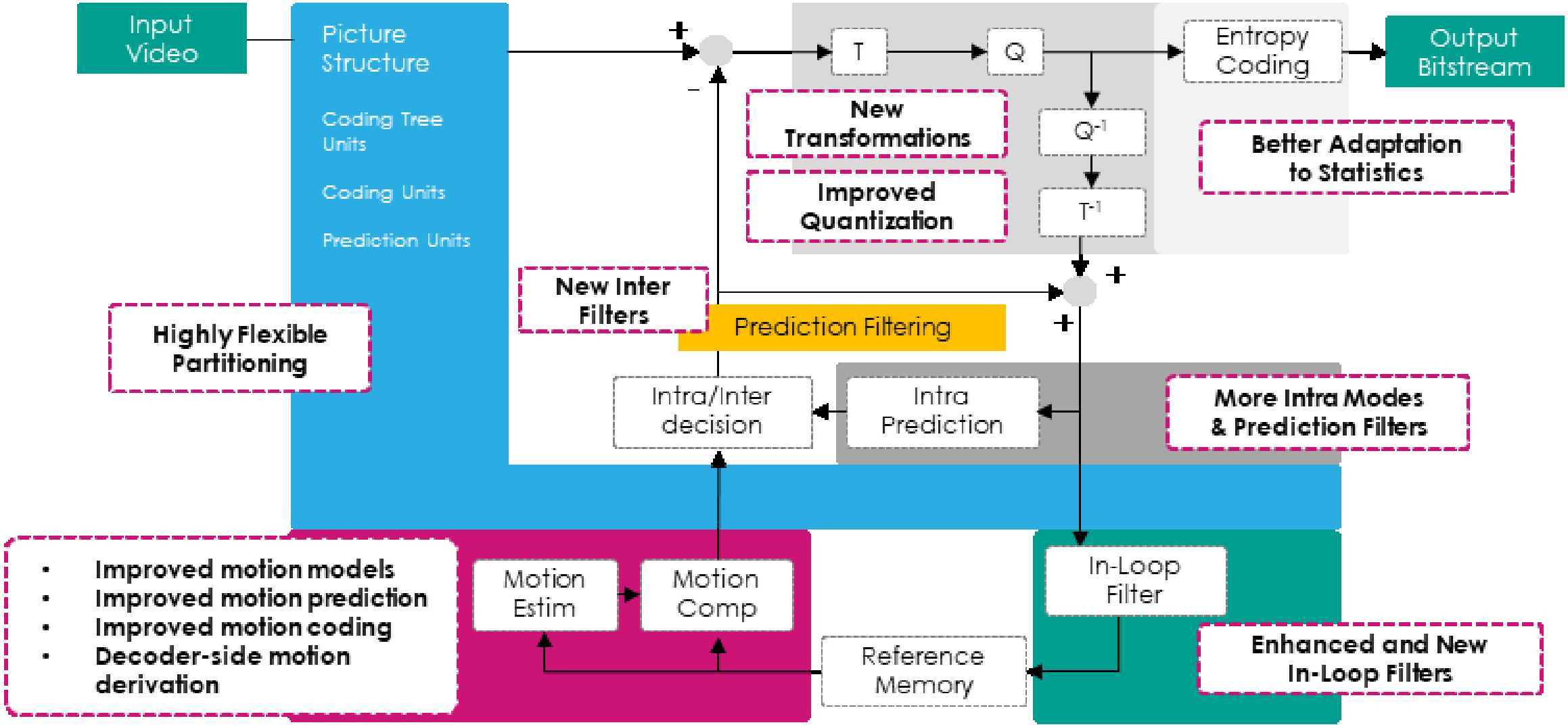
© 2020 InterDigital, Inc. All Rights Reserved.

Video Coding Standards

Standard	Organization	Application	Licensing Condition	When
Versatile Video Coding (VVC) **MPEG-I Part 3 (IEC/ISO 23090-3) /*H.266*** **ISO/IEC 23002-7 Versatile SEI / H.274**	ITU-T, ISO/IEC (JVET: MPEG & VCEG)	2D planar video: VOD, Broadcasting, Conferencing, Screen Content; 360° video	ISO/IEC,ITU-T IPR Policy (FRAND)	2020
Essential Video Coding (EVC) / **MPEG-5 Part 1 (IEC/ISO 23094-1)**	ISO/IEC (MPEG : Samsung, Huawei, Qualcomm, Tencent)	2D planar video: VOD, Broadcasting, Conferencing, Screen Content	ISO/IEC,ITU-T IPR Policy • Baseline is RAND RF • Main is FRAND	2020
Low Complexity Essential Video Coding (LC-EVC) / **MPEG-5 Part 2 (IEC/ISO 23094-2)**	ISO/IEC (MPEG : V-nova)	2D planar video: where a base codec exists and low complexity is key	ISO/IEC,ITU-T IPR Policy	2020 (probably delayed to 2021)
AOMedia Video 1 (AV1)	AOM (Google, Cisco, Amazon, Netflix …)	2D planar video: VOD, Conferencing, Screen Content	Royalty-free	2018
Audio Video Coding Standard 3 (AVS3)	China AVS workgroup (Huawei, Tencent, Bytedance …)	2D planar video: VOD, Broadcasting, Conferencing, Screen Content; 360° video	AVS IPR Policy: • RAND RF • POOL • RAND	2020

© 2020 InterDigital, Inc. All Rights Reserved.

Video coding improvements wrt HEVC



© 2020 InterDigital, Inc. All Rights Reserved.
interdigital

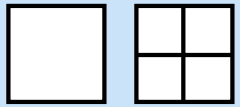
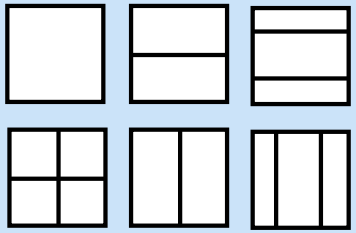
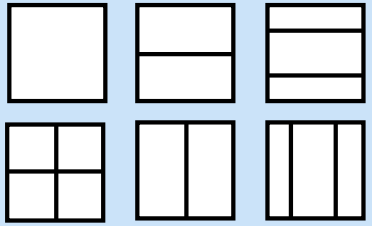
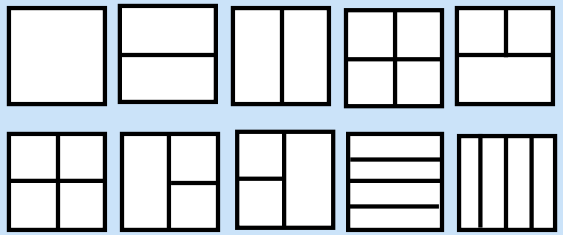
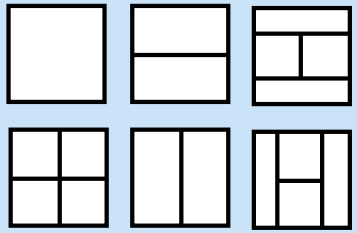
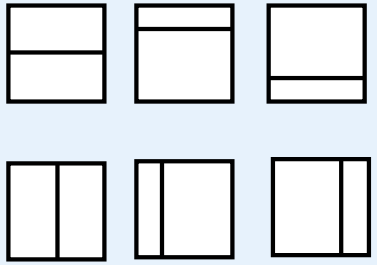
Partitioning

Category	HEVC	VVC	AV1	EVC	AVS3
Picture partitioning	• Slices • Tiles	• Sub-pictures • Slices • Tiles	• Tiles • Tile Group	• Slices • Rectangular slices with tiles	• Patch: similar as the rectangular slices, but with unified width/height
Block Partitioning	• CTU: 16x16, 32x32 or 64x64 • CU: QT*, 8x8 to 64x64	• CTU: 32x32, 64x64 or 128x128 • CU: QT, BT*, TT* - 4x4 to 128x128 • Dual-tree in I Slices	• Superblock: 10 ways split - 4x4 to 128x128	• CTU: 32x32, 64x64 or 128x128 • QT + BT + TT	• CTU: 32x32, 64x64 or 128x128 • QT + BT + EQT*
TU Partitioning	• Recursive QT, 32x32-> 4x4	• QT (Tiling) • BT, ABT* (SBT*) • BT, 4-split (ISP*) • 4x4	• 2-level recursive quad-tree transform size partition for inter Luma	• BT, ABT* (For inter blocks using Adaptive Transform Selection)	• DT*
PU Partitioning	• Non-recursive partitioning of a CU (BT, AMP*), 4x4 intra, 8x4 inter	• GEO	• Wedges		
Block Coding Order				• SUCO*	

*QT: Quad-Tree
BT: Binary Tree
TT: Ternary Tree
EQT: Extended Quad-Tree
ABT: Asymmetric Binary Tree
SBT: Sub-Block Transform
ISP: Intra Sub-Partitions
DT: Derived Tree
AMP: Asymmetric Motion Partition
SUCO: Split Unit Coding Order

© 2020 InterDigital, Inc. All Rights Reserved.

Partitioning

	HEVC	EVC	VVC	AV1	AVS3
CU or PU					
PU only			64 geometric positions ...	16 wedge partitions ...	

© 2020 InterDigital, Inc. All Rights Reserved.

Intra Prediction

Category	HEVC	VVC	AV1	EVC	AVS3
Prediction model	• 33 angles + 2 (planar, DC) prediction modes	• 65 regular angles + 28 wide angles + 2 prediction modes • MIP* • MRL*	• 56 angles + DC + 3 smooth predictors + PAETH predictor	• 30 angles + 2 • Bilinear	• 62 angles + 2 • Bilinear
Prediction coding	• 3 MPM*	• 6 MPM		• 2 MPM + 2nd 8 MPM	• 2 MPM
Chroma		• CCLM*	• CFL*		• TSCPM*
Filtering	• Reference samples filtering (gaussian/no filter) • Linear interpolation of reference samples • boundary filtering of the prediction	• reference samples filtering (gaussian/no filter) • cubic/gaussian/linear interpolation of reference samples • PDPC*	• Recursive filtering-based intra predictor		• IPF*

*MIP: Matrix-based Intra Prediction

MRL: Multi-Reference Lines intra prediction

MPM: Most Probable Modes

CCLM: Cross-Component Linear Model

CFL: Chroma From Luma

TSCPM: Two-Step Cross-component Prediction Mode

PDPC: Position-Dependent Prediction Combination

IPF: Intra Prediction Filter

© 2020 InterDigital, Inc. All Rights Reserved.

Inter Prediction

Category	HEVC	VVC	AV1	EVC	AVS3
Motion info. Prediction	• Spatial • Temporal • Combined motion info. • Zero motion	• Spatial • Temporal • Pairwise • HMVP* • Reconstructed affine model • SbTMVP*	• Spatial • Extended spatial • Temporal • Average • Zero motion	• Spatial • Temporal • HMVP • Combined motion info. • Zero motion	• Spatial • Temporal • HMVP
Prediction coding type	• Merge • AMVP*	• Merge • AMVP • MMVD* • AMVR* • GEO* • SMVD*	• Merge • AMVP	• Merge • AMVP • MMVD • AMVR	• Merge • AMVP • AMVR • SMVD • UMVE* • EMVR*
Motion model	• Translational	• Translational • Affine (4 or 6 param.) • SbTMVP • DMVR* • GEO	• Translational • Local affine (4 or 6 param.) • Global affine • Wedge prediction	• Translational • Affine (4 or 6 param.) • DMVR	• Translational • Affine (4 or 6 param.) • DMVR

*HMVP: History-based MV Prediction

SbTMVP: Sub-block based Temporal MV Prediction

AMVP: Adaptive MV Prediction

MMVD: Merge with MV Difference

AMVR: (Locally) Adaptive MV Resolution

GEO: Geometric Partitioning Mode

SMVD: Symmetric MV Difference

UMVE: Ultimate MV Express

EMVR: Combined HMVP&AMVR

DMVR: Decoder-side MV Refinement

© 2020 InterDigital, Inc. All Rights Reserved.

Inter Prediction

Category	HEVC	VVC	AV1	EVC	AVS3
Prediction blending	• Unidirectional • Bidirectional average	• Unidirectional • Bidirectional average • BCW* • CIIP* • RPR*	• Difference-modulated mask prediction • Frame distance-based compound • Compound inter-intra prediction • Super-resolution (~RPR)	• Unidirectional • Bidirectional average	
Prediction enhancement		• BDOF* • PROF* • SIF*	• OBMC* • Independent switchable interpolation filter	• 2 Interpolation filters (Luma: 1, Chroma: 1)	

*BCW: Bi-prediction with CU based weighting

CIIP: Combined Inter/Intra Prediction

BDOF: Bi-Directional Optical Flow

PROF: Prediction Refinement with Optical Flow

SIF: Switchable (motion) Interpolation Filter

OBMC: Overlapped Block Motion Compensation

RPR: Reference Picture Resampling

© 2020 InterDigital, Inc. All Rights Reserved.

Residual Coding

Category	HEVC	VVC	AV1	EVC	AVS3
Transform	• DCT2 square (4x4 to 32x32) • Intra 4x4: DST4	• DCT2 (2x2 to 64x64, zero-out>32) • MTS* (up to 32x32, zero-out>16, DCT2, DST7, DCT8) • SBT* for inter • LFNST*	• DCT2 • ADST • Flipped ADST • Identity • 2-level recursive quad-tree transform size partition for inter Luma	• DCT2 • ATS* (up to 32x32, DCT2, DST7, DCT8)	• DCT2 • PBT*
Quantization	• QP: 0..51 • Normal Scaling	• QP: 0..51 • Normal Scaling • DQ*	• Adaptive scaling	• QP: 0..51 • Normal Scaling	
Residual coding	• ACT* • Coefficient coding • Sign hiding	• ACT • JCCR* • Coefficient coding • Coefficient coding for TS*/BDPCM* • Parity-based binarization for DQ • Sign hiding	• Level map coefficient coding	• ADCC*	• SRCC*

*MTS: Multiple Type Transforms

SBT: Sub-Block Transforms

LFNST: Low-Frequency Non-Separable Transform

ATS: Adaptive Transform Selection

PBT: Position Based Transform

DQ: Dependent Quantization

ACT: Adaptive Colour Transform (4:4:4)

JCCR: Joint Coding of Chroma Residual

TS: Transform Skip

BDPCM: Block-wise DPCM

ADCC: Advanced Coefficient Coding

SRCC: Scan Region based Coefficient Coding

© 2020 InterDigital, Inc. All Rights Reserved.

Entropy Coding & Loop Filters & SCC

Category	HEVC	VVC	AV1	EVC	AVS3
Entropy coding	• CABAC*	• CABAC • Multi-hypothesis probability estimation • Context-adaptive window size	• Multi-symbol arithmetic coder	• CABAC	• CABAC
Loop filters	• Deblocking filter • SAO*	• Long tap deblocking filter • SAO • ALF* • CC-ALF* • LMCS*	• Deblocking filter • CDEF* • Loop restoration	• Deblocking filter • ALF* • HTDF*	• Deblocking filter • SAO • ALF
SCC	• Palette mode • IBC* • ACT*	• Palette mode • IBC • ACT • Transform Skip • BDPCM*	• Palette mode • IBC	• IBC	• IBC
Post-filter (HDR)				• DRA*	

*CABAC: Context Adative Binary Arithmetic Coding

SAO: Sample Adaptive Offset

ALF: Adaptive Loop Filter

CC-ALF: Cross-component ALF

LMCS: Luma Mapping with Chroma Scaling

CDEF: Constrained Directional Enhancement Filter

HTDF: Hadamard Transform Domain Filter

IBC: Intra Block Copy

ACT: Adaptive Chroma Transform

BDPCM: Block-wise DPCM

DRA: Dynamic Range Adjustment

© 2020 InterDigital, Inc. All Rights Reserved.

Other tools

Category	HEVC	VVC	AV1	EVC	AVS3
Scalability	• Spatial • Quality • temporal	Thanks to RPR: • Spatial • Quality • Temporal			
High Level Syntax		• GDR: Gradual Decoding Refresh • BEAM: Bit Stream Extraction and Merge			

© 2020 InterDigital, Inc. All Rights Reserved.

Test conditions

- **Comparison according to common Test conditions (CTC)**
  - Random access case of JVET
  - 6 UHD and 5 HD sequences
  - 8K (non-CTC)
- **VVC, EVC and HEVC have reference softwares giving the state of the art:**
  - VVC: **VTM9.3** (VVC Test Model) in Main10 profile
  - EVC: **ETM6.1** (Test Model) in Main profile
  - HEVC: **HM16.18** (HEVC Model) in Main 10 profile
- **AV1 reference software:**
  - **Libaom2.0** in cpu0 mode (maximum quality)
- **AV1 configuration close to CTC used for VTM, ETM, HM**
- **AVS3 not part of the tests**

© 2020 InterDigital, Inc. All Rights Reserved.

AV1 configuration

- **Configuration to match as much as possible JVET-CTC**
- **Double-pass to get the best performance**
  - Normal use case of AV1
  - Libaom not performant (not maintained?) in single pass
  - Impact on short test sequences should be low
- **Configuration Parameters:**
  - GOF (Group of Frames)
  - Hierarchical GOF
  - Constant Quantization (cq-level)
  - Fixed quantization per image (delta-q=0)
  - Maximum quality (longest run time) (cpu0)

© 2020 InterDigital, Inc. All Rights Reserved.

Broadcast scenario

- **HM, VTM, and ETM:**
  - Intra period ~= 1 second
  - GOP size = 16
  - Hierarchical GOP
  - Fixed Quantization Parameter per picture type
- **AV1:**
  - Double pass encoding
  - Intra period ~= 1 second
  - GOF size = 16
  - Hierarchical GOP
  - Fixed Quantization Parameter per picture type
    - deltaQ = 0

© 2020 InterDigital, Inc. All Rights Reserved.

Streaming scenario

- **HM, VTM, and ETM:**
  - Intra period ~= **2** second (*)
  - GOP size = 16
  - Hierarchical GOP
  - Fixed Quantization Parameter per picture type
- **AV1:**
  - Double pass encoding
  - Intra period ~= **2** second (*)
  - GOF size = 16
  - Hierarchical GOP
  - Fixed Quantization Parameter per picture type
    - deltaQ = 0

(*) DASH recommends 2s segments

© 2020 InterDigital, Inc. All Rights Reserved.

Metrics

- **PSNR (Peak Signal to Noise ratio)**
  - $PSNR = 10\log_{10}(Max^2 / MSE)$
  - $PSNR_{YUV} = (6 \times PSNR_Y + PSNR_U + PSNR_V) / 8$
  - $PSNR_Y$

- **VMAF (Video Multi-method Assessment Fusion)**
  - Score based on several objective metrics
  - Machine Learning to introduce subjective aspects (trained models)

- **MS-SSIM (Multi-Scale Structural Similarity)**
  - Measure similarity of structural information
  - Some perceptual information (masking effects…)

© 2020 InterDigital, Inc. All Rights Reserved.

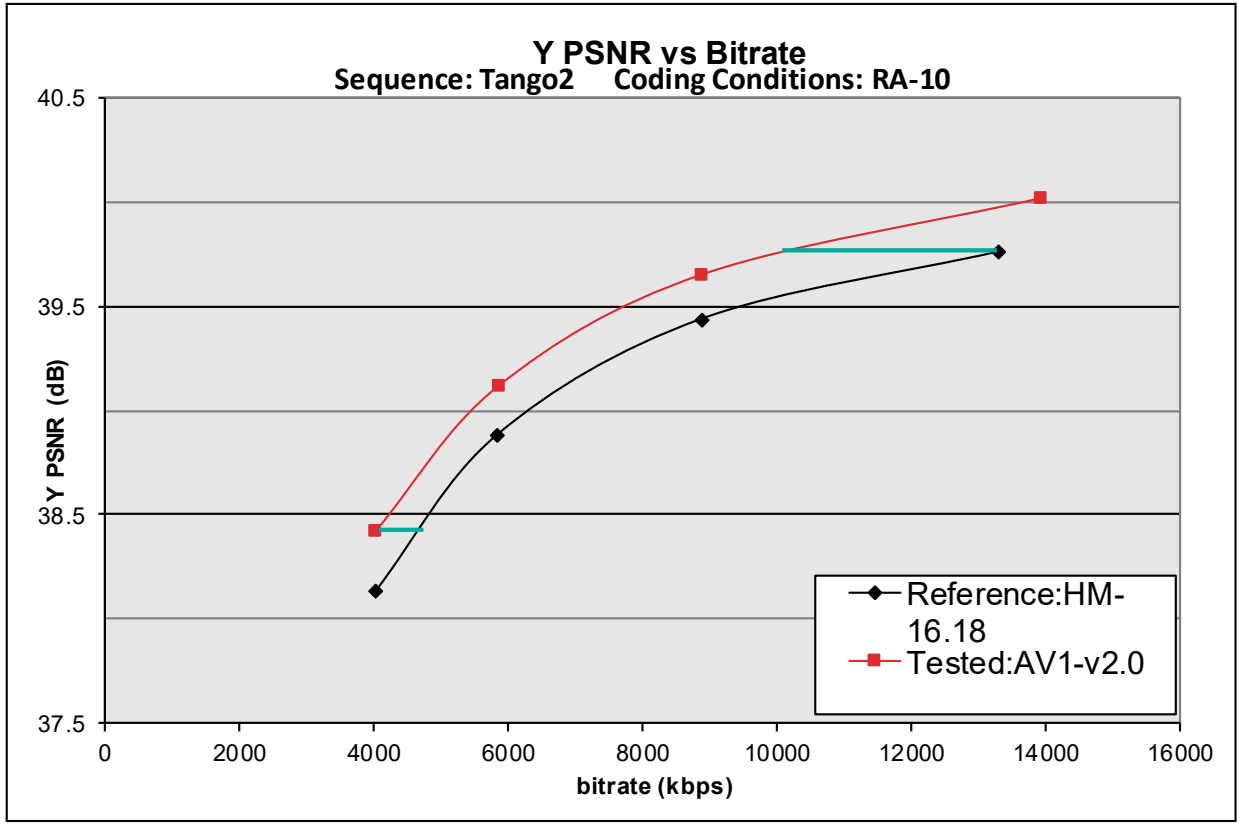
Performance measure

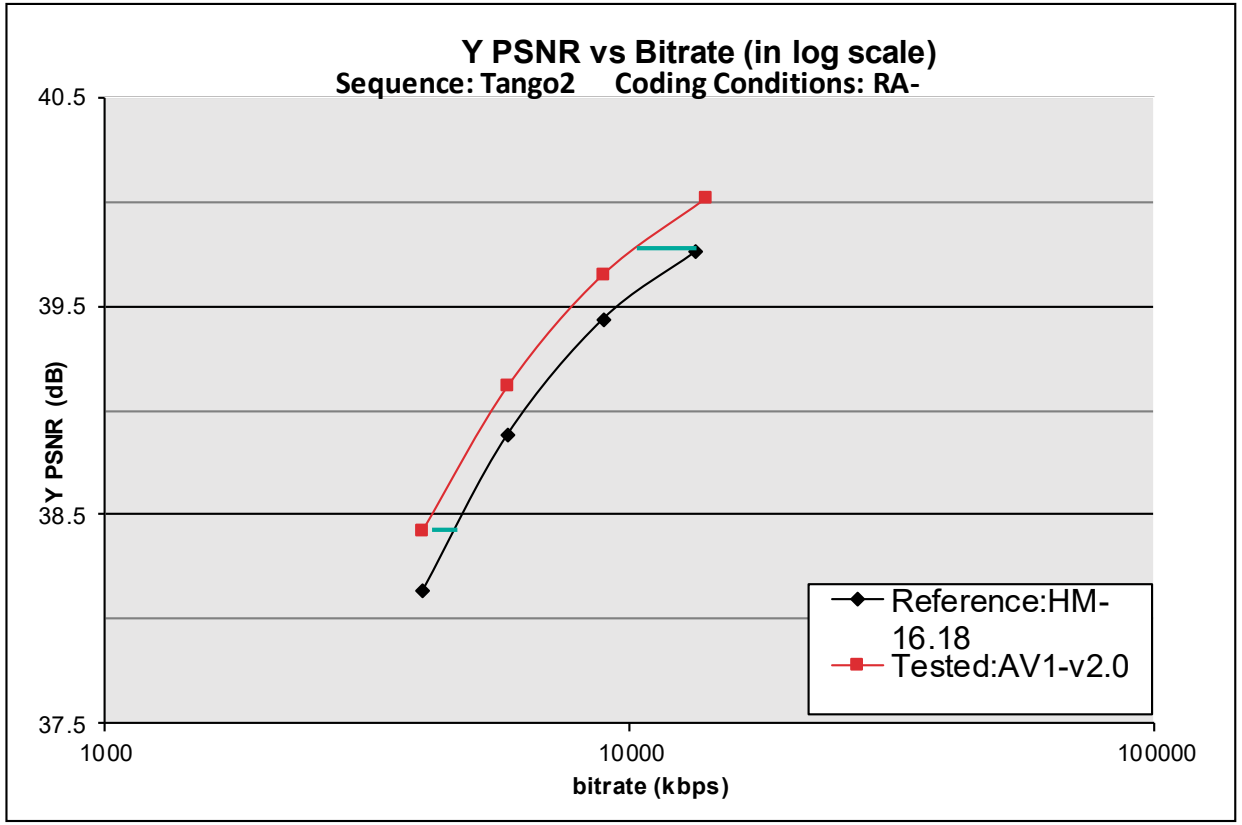
- **Based on ITU-T HSTP-VID-WPOM (07/2020):**
  - "Working practices using objective metrics for evaluation of video coding efficiency experiments".
- **Bjøntegaard Delta-Rate" (BD-rate) metric**
- **Measure a gain in bit rate at constant quality metric:**
  - Versus a reference: here the **HM**
  - Percentage value (negative means a gain)
  - Quality metrics:
    - $PSNR_{YUV}$
    - $PSNR_{Y}$
    - VMAF
    - MS-SSIM
  - Bit rates determined from AV1

© 2020 InterDigital, Inc. All Rights Reserved.

Performance measure

- Measure the area between the 2 curves
- Divide by the metric range





© 2020 InterDigital, Inc. All Rights Reserved.

Broadcast scenario

$PSNR_{YUV}$ BD-rate Vs HM

Broadcast PSNR$_{YUV}$	VTM9.3	libaom 2.0	ETM6.1
UHD	-41.2%	-23.0%	-27.9%
HD	-37.8%	-23.7%	-23.4%
Overall	-39.7%	-23.3%	-25.8%
Diff vs VTM	0.0%	16.3%	13.8%

$PSNR_Y$ BD-rate Vs HM

Broadcast PSNR$_Y$	VTM9.3	libaom 2.0	ETM6.1
UHD	-42.1%	-19.9%	-28.6%
HD	-35.8%	-20.6%	-23.1%
Overall	-39.2%	-20.2%	-26.1%
Diff vs VTM	0.0%	19.0%	13.1%

VMAF BD-rate Vs HM

Broadcast VMAF	VTM9.3	libaom 2.0	ETM6.1
UHD	-44.9%	-24.3%	-33.5%
HD	-41.6%	-28.1%	-30.4%
Overall	-43.4%	-26.0%	-32.1%
Diff vs VTM	0.0%	17.3%	11.3%

MS-SSIM BD-rate Vs HM

Broadcast MS-SSIM	VTM9.3	libaom 2.0	ETM6.1
UHD	-41.1%	-19.8%	-29.4%
HD	-34.1%	-18.5%	-23.6%
Overall	-37.9%	-19.2%	-26.8%
Diff vs VTM	0.0%	18.7%	11.1%

In UHD: VTM **> 41.2%** ; Libaom **> 19.8%** ; ETM **> 27.9%** on the four metrics

© 2020 InterDigital, Inc. All Rights Reserved.

Streaming scenario

$PSNR_{YUV}$ BD-rate Vs HM

Streaming $PSNR_{YUV}$	VTM9.3	libaom 2.0	ETM6.1
UHD	-40.7%	-24.9%	-27.2%
HD	-35.4%	-25.9%	-21.7%
Overall	-38.3%	-25.4%	-24.7%
Diff vs VTM	0.0%	12.9%	13.6%

$PSNR_Y$ BD-rate Vs HM

Streaming $PSNR_Y$	VTM9.3	libaom 2.0	ETM6.1
UHD	-41.7%	-21.8%	-27.9%
HD	-33.6%	-22.4%	-21.5%
Overall	-38.0%	-22.1%	-25.0%
Diff vs VTM	0.0%	15.9%	13.0%

VMAF BD-rate Vs HM

Streaming VMAF	VTM9.3	libaom 2.0	ETM6.1
UHD	-44.6%	-25.4%	-33.2%
HD	-39.5%	-29.8%	-28.9%
Overall	-42.3%	-27.4%	-31.3%
Diff vs VTM	0.0%	14.9%	11.0%

MS-SSIM BD-rate Vs HM

Streaming MS-SSIM	VTM9.3	libaom 2.0	ETM6.1
UHD	-40.8%	-21.8%	-28.4%
HD	-31.3%	-21.1%	-21.4%
Overall	-36.5%	-21.5%	-25.2%
Diff vs VTM	0.0%	15.0%	11.3%

In UHD: VTM **> 40.7%** ; Libaom **> 21.8%** ; ETM **> 27.2%** on the four metrics

© 2020 InterDigital, Inc. All Rights Reserved.

Broadcast scenario

- **Reference softwares evolution (BD-rate Vs HM):**
  - Libaom shows performance improvements Vs previous versions
  - VTM and ETM are stable

Broadcast UHD	$PSNR_{YUV}$	$PSNR_{Y}$	VMAF	MS-SSIM	EncT	DecT
VTM9.3	-41.2%	-42.1%	-44.9%	-41.1%	1142%	194%
libaom2.0	-23.0%	-19.9%	-24.3%	-19.8%	621%	74%
ETM6.1	-27.9%	-28.6%	-33.5%	-29.4%	605%	139%

Broadcast UHD	$PSNR_{YUV}$	$PSNR_{Y}$	VMAF	MS-SSIM	EncT	DecT
VTM8	-41.9%	-38.7%	-42.8%	-39.5%	1308%	162%
libaom1.0	-18.0%	-18.0%	-19.6%	-14.5%	497%	76%
ETM4.1	-28.4%	-28.2%	-33.9%	-29.9%	669%	167%

© 2020 InterDigital, Inc. All Rights Reserved.

Streaming scenario

- **Reference softwares evolution (BD-rate Vs HM):**
  - Libaom shows performance improvements Vs previous versions
  - VTM and ETM are stable

Streaming UHD	$PSNR_{YUV}$	$PSNR_{Y}$	VMAF	MS-SSIM	EncT	DecT
VTM9.3	-40.7%	-41.7%	-44.6%	-40.8%	1173%	193%
libaom2.0	-24.9%	-21.8%	-25.4%	-21.8%	627%	75%
ETM6.1	-27.2%	-27.9%	-33.2%	-28.4%	618%	139%

Streaming UHD	$PSNR_{YUV}$	$PSNR_{Y}$	VMAF	MS-SSIM	EncT	DecT
VTM8	-41.2%	-38.7%	-41.6%	-38.0%	1283%	201%
libaom1.0	-20.5%	-18.0%	-21.5%	-17.0%	515%	77%
ETM4.1	-27.7%	-28.2%	-33.3%	-28.8%	680%	167%

© 2020 InterDigital, Inc. All Rights Reserved.

Results on 8K

- **8K sequences provided by HHI (Fraunhofer)**
  - RED DSMC2 camera in 7680x4320 60Hz 10 bits 4:2:0 (10s)
  - Log3G10 transfer function
  - REDWideGamutRGB color space
    - BodeMuseum
    - NeptuneFountain2
    - NeptuneFountain3
    - OberbaumSpree
- **Indeed non-JVET CTC sequences**
- **Random access (1s) only**

© 2020 InterDigital, Inc. All Rights Reserved.

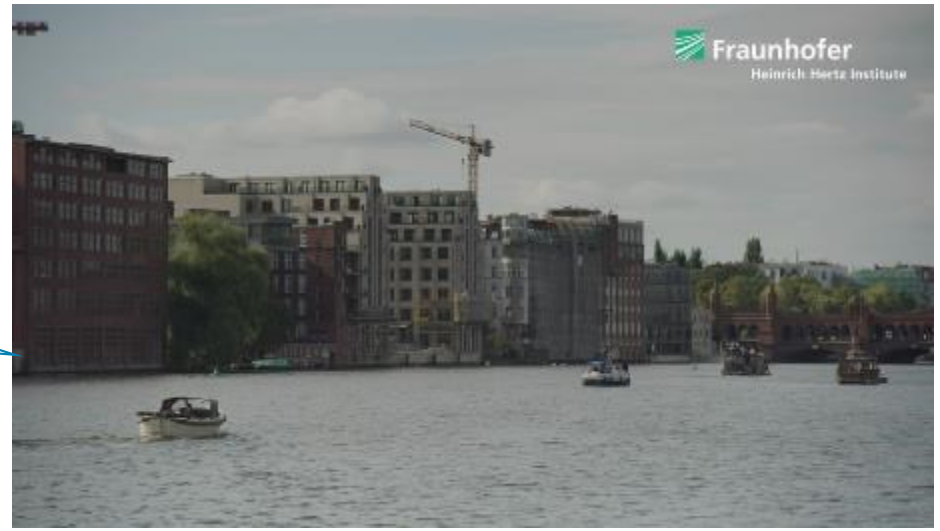
8K sequences

BodeMuseum

Neptune Fountain2

Neptune Fountain2

OberBaum Spree



© 2020 InterDigital, Inc. All Rights Reserved.
interdigital

Results on 8K

Broadcast 8K	$PSNR_{YUV}$	$PSNR_{Y}$	VMAF	MS-SSIM
VTM9.3	-31.2%	-29.2%	-33.8%	-27.9%
libaom2.0	-17.3%	-13.8%	-20.7%	-10.9%
ETM6.1	-13.2%	-11.2%	-18.2%	-9.6%

Sequence	BD-rate $PSNR_{YUV}$		
	VTM	AV1	ETM
BodeMuseum	-31.7%	-10.0%	-9.5%
NeptuneFountain2	-29.9%	-21.9%	-23.0%
NeptuneFountain3	-27.8%	-16.2%	-15.9%
OberbaumSpree	-35.5%	-21.0%	-4.3%
All	-31.2%	-17.3%	-13.2%

- **VTM performances consistent with:**
  - Ateme IBC conference 2020:
    - -33% on 8K non-CTC
  - Interdigital IBC conference 2020 VVC per tool evaluation
    - -35% BD-rate gain on 4K non-CTC sequences
  - NHK JVET contribution Oct 2019: -30% BD-rate gain for HLG 8K
- **Libaom2.0 reduced the gap with VTM (18% to 14%)**
- **ETM gap with VTM is larger than in UHD CTC sequences (13% to 18%)**

© 2020 InterDigital, Inc. All Rights Reserved.

Conclusion

- **Results consistent with published information**
- **On JVET CTC sequences, for broadcast, in UHD, wrt HM**
  - VTM: **> 41.2%** on the four metrics
  - Libaom: **> 19.8%** on the four metrics (advantage of 2-pass)
  - ETM: **> 27.9%** on the four metrics
- **The results in 8K are consistent for non-CTC sequences**
  - Except for ETM
- **Deployment of new codecs will depend on the capabilities to meet the technical requirements of new immersive services**
- **But also on business conditions**

© 2020 InterDigital, Inc. All Rights Reserved.

Thank you

Acknowledgement to the InterDigital video coding team

© 2020 InterDigital, Inc. All Rights Reserved.

interdigital