HEVC Subjective Video Quality Test Results

T. K. Tan	NTT DOCOMO, Japan
M. Mrak	BBC, UK
R. Weerakkody	BBC, UK
N. Ramzan	University of West of Scotland, UK
V. Baroncini	Fondazione Ugo Bordoni, Italy
G. J. Sullivan	Microsoft Corporation, USA
J.-R. Ohm	RWTH Aachen University, Germany
K. D. McCann	Zetacast, UK

Overview of Presentation

Video coding standards

Performance measurements

HEVC verification methods

HEVC verification test results

Conclusions

Major Video Coding Standards

Mid 1990s: MPEG-2

Mid 2000s: H.264/AVC

Mid 2010s: HEVC

These video coding standards are the joint work of the same two bodies

- ISO/IEC Moving Picture Experts Group (MPEG)
- ITU-T Video Coding Experts Group (VCEG)
- Most recently working on High Efficiency Video Coding (HEVC) as Joint Collaborative Team on Video Coding (JCT-VC)

HEVC version 1 was completed in January 2013

- published by ISO/IEC as ISO/IEC 23008-2
- published by ITU-T as H.265

Comparison of HEVC and H.264/AVC

	HEVC	H.264/AVC
Coding Tree Unit	64x64, 32x32, 16x16 CTU	16x16 macroblock
Coding Unit	64x64, 32x32, 16x16, 8x8 CU	16x16 macroblock
Prediction Unit	square, symmetric rectangular, asymmetric rectangular PU	square, symmetric rectangular
Transform Unit	32x32, 16x16, 8x8, 4x4 TU	8x8, 4x4 transforms
Intra prediction	33 directional modes, DC, planar	8 directional modes, DC, planar
Motion prediction	multi-candidate MV prediction with spatial and temporal region merging	spatial median or temporal co-located motion vector prediction
Luma interpolation	¼ pixel 7-tap, ½ pixel 8-tap	½ pixel 6-tap + ¼ pixel bilinear
Chroma interpolation	4-tap	bilinear
Entropy coding	CABAC	CABAC, CAVLC
Loop filtering	deblocking filter, sample adaptive offset	deblocking filter
Parallelism	tiles, wavefronts, slices	slices

Performance Measurements

Two different types of measurement of compression performance

- Objective measurement, e.g. Peak Signal-to-Noise Ratio (PSNR)
  - Easy to calculate, but only indicative of actual quality perceived by viewers
- Subjective evaluation
  - Mean Opinion Score (MOS) from human test subjects
  - Expensive and time-consuming formal subjective tests
  - Requires a large number of test subjects to give good confidence intervals

Compression doesn’t depend only on the coding standard

- Encoder implementation, video test sequences, etc.

JCT-VC Verification Testing of HEVC

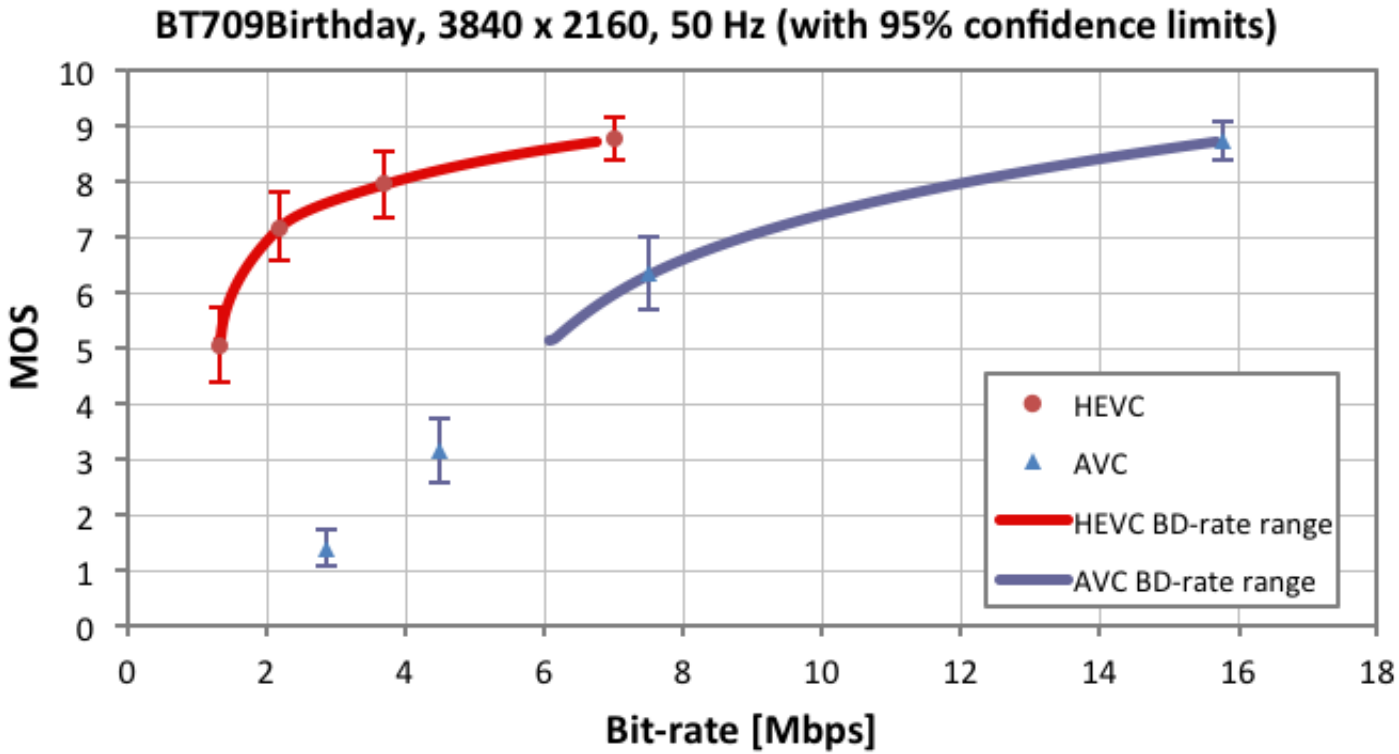
Verification testing compared HEVC and H.264/AVC standards

- Encoding used the two standard test models
  - HEVC Main profile using HM12.1
  - H.264/AVC High profile using JM18.5
- 20 test sequences covering four resolutions
  - 480p, 720p, 1080p, UHD-1 (2160p)

Each sequence was evaluated subjectively at four bit rates

- Degradation Category Rating (DCR) method
- Rated using quality scale from 0 to 10

Example of UHD-1 Sequence



Verification Test Results

Subjective test results require careful analysis

- Mean Opinion Score (MOS) calculated for each test point
- Opinions of viewers who gave less than 0.75 correlation with general opinion were regarded as unreliable and were omitted
- Revised MOS were then plotted with 95% confidence intervals

Each HEVC test point categorised

- Compared to H.264/AVC points with overlapping confidence intervals

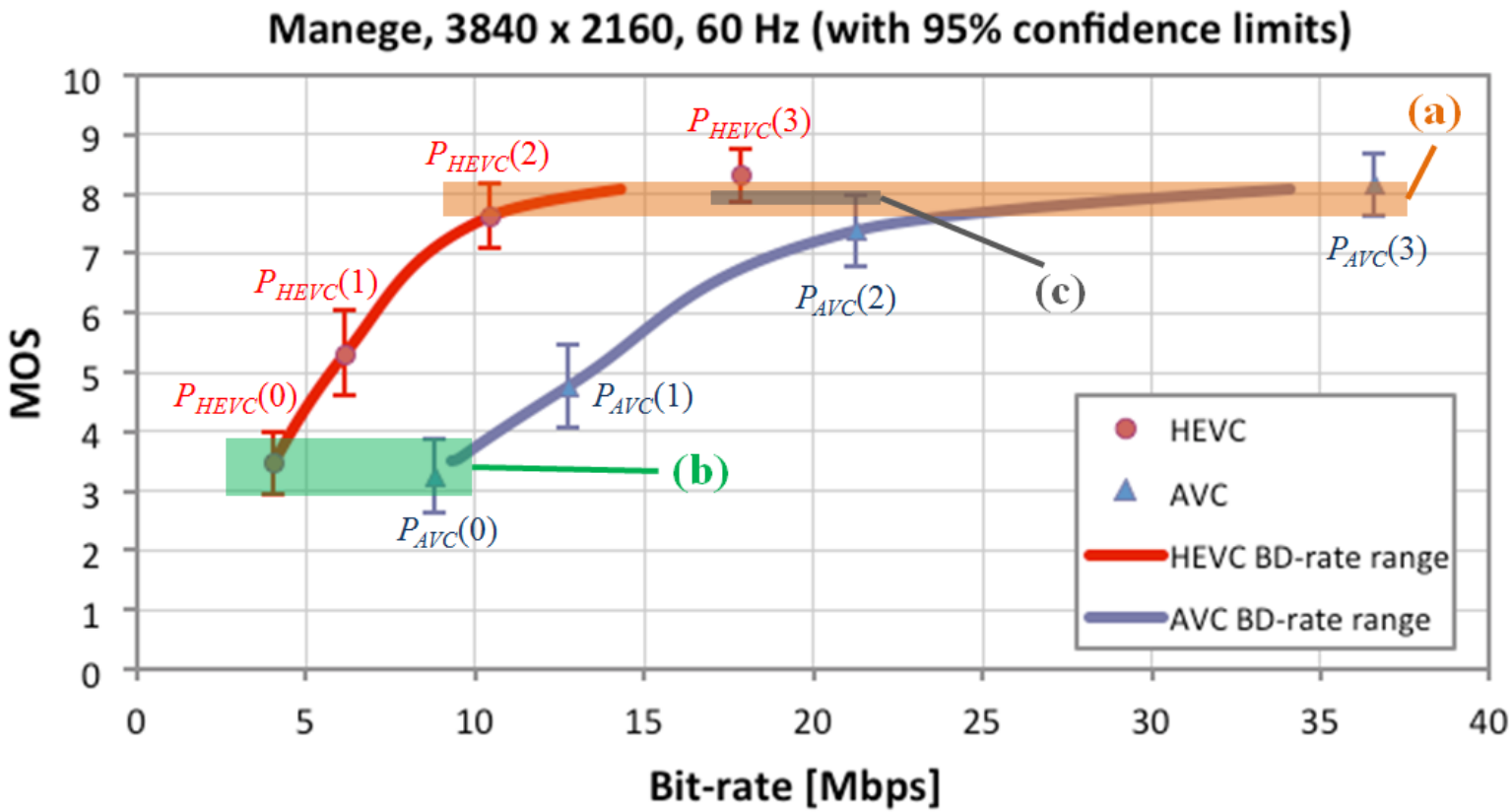
Bit rate required for HEVC was

(a) Less than half that of H.264/AVC for 41% of the comparisons

(b) About half that of H.264/AVC for 45% of the comparisons

(c) More than half that of H.264/AVC for 14% of the comparisons

Example of test point categories



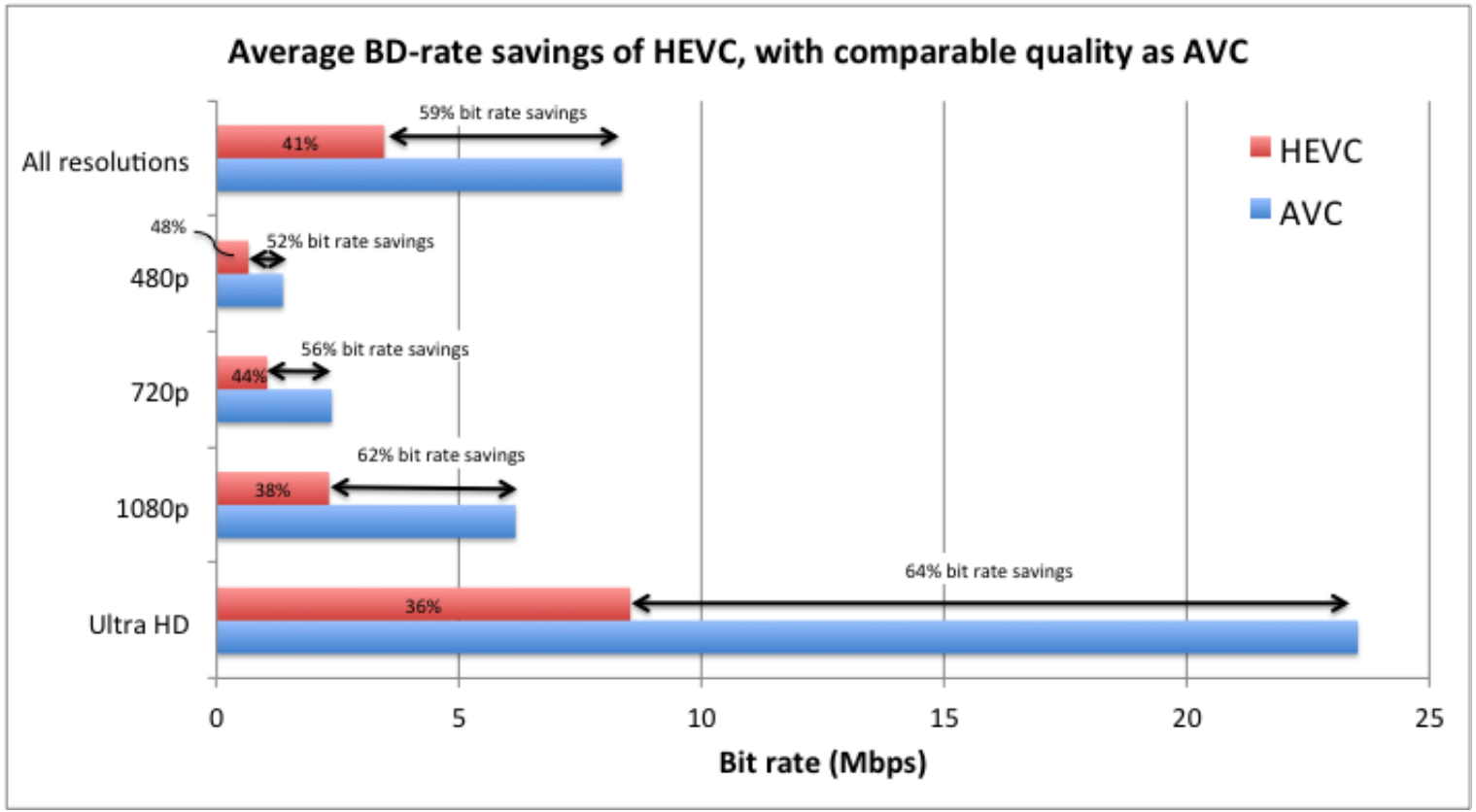
Estimated efficiency improvement

Resolution	Sequence	MOS BD-rate
UHD-1 (2160p)	BT709Birthday	−75%
	Book	−66%
	HomelessSleeping	*
	Manege	−56%
	Traffic	−58%
1080p	JohnnyLobby (LD)	−70%
	Calendar	−52%
	SVT15	−69%
	sedofCropped	−53%
	UnderBoat1	−68%
720p	ThreePeople (LD)	−48%
	BT709Parakeets	−66%
	QuarterBackSneak	−58%
	SVT01a	−73%
	SVT04a	−36%
480p	Cubicle (LD)	−45%
	Anemone	−42%
	BT709BirthdayFlash	−49%
	Ducks	−72%
	WheelAndCalender	*
Average		−59%

Efficiency calculation used Bjøntegaard Delta rate on the MOS scores

- Standard statistical tool used for objective measurements
- Applied to MOS scores using piecewise cubic interpolation
- No valid calculation possible on two sequences that didn't give smooth curves

Average bit rate savings



Conclusions

HEVC required no more than half bit rate of H.264/AVC in 86% of cases

- Implies that HEVC target of doubling the compression efficiency has been met or exceeded

Bjøntegaard Delta calculations indicate greater bit rate savings at the higher resolutions

- 52% for 480p
- 56% for 720p
- 62% for 1080p
- 64% for UHD-1 (2160p)

Subjective performance confirmed to be better than would be predicted from the objective test results

Thank you for your attention

and thanks to all organisations and individuals who contributed to the HEVC verification tests