The Arrival of the High Efficiency Video Coding Standard (HEVC)

Gary J. Sullivan
Co-Chair JCT-VC, Co-Chair JCT-3V, Chair VCEG, Co-Chair MPEG Video, Video & Image Technology Architect, Microsoft  <garysull@microsoft.com>
20 March 2013

Presentation for Data Compression Conference (DCC 2013)

H.264/MPEG-4 Advanced Video Coding (AVC): The basic idea (2003)
- Compress digital video content
- Twice as much as you did before
- With the same video quality, e.g. as MPEG-2 or H.263
- Or get higher quality with the same number of bits (or a combo)
- Example: higher quality may mean higher resolution, e.g. HD
- And better adaptation to applications and network environments
  - Unfortunately, with substantially higher computing requirements and memory requirements for both encoders and decoders
    - But this time the decoder is not so tough (~1.5x)
    - And the memory increase is not so much
    - And the parallelism opportunities are better (throughout)

High Efficiency Video Coding (HEVC): The basic idea (2013)
- Compress digital video content
- Twice as much as you did before
- With the same video quality, e.g. as AVC
- Or get higher quality with the same number of bits (or a combo)
- Example: higher quality may mean higher resolution, e.g. Ultra-HD
- And better adaptation to applications and network environments
- Unfortunately, with substantially higher computing requirements and memory requirements for both encoders and decoders
  - But this time the decoder is not so tough (~1.5x)
  - And the memory increase is not so much
  - And the parallelism opportunities are better (throughout)

Chronology of International Video Coding Standards

Video Coding Standards Organizations
- JVT = “Joint Video Team” collaborative team of MPEG & VCEG
- SMPTE (Society of Motion Picture and Television Engineers)
- New: JCT-VC = “Joint Collaborative Team on Video Coding” team of MPEG & VCEG, continuing the collaborative relationship for a new project (established January 2010)
- New: JCT-3V = “Joint Collaborative Team on 3D Video” extension development team of MPEG & VCEG for 3D (est. May 2012)

The Scope of Video Coding Standardization
- Only the Syntax and Decoder are standardized:
  - Permits optimization beyond the obvious
  - Permits complexity reduction for implementability
  - Provides some capability, but no guarantee of Quality
HEVC and the JCT-VC Partnership

- Initial groundwork in VCEG and MPEG
- "Key Technical Areas (KTA)" study and software in VCEG
- "Call for Evidence" in MPEG
- Agreement in January 2010 to form new team VCEG-AM90 / N11112
- Joint Call for Proposals on Video Coding Technology issued January 2010 VCEG-AM91 / WG 11 N11113
- Joint Collaborative Team (JCT) on Video Coding (JCT-VC)
- Chairs: Gary Sullivan (Microsoft) & Jens-Rainer Ohm (RWTH Aachen Univ.)
- Project name High Efficiency Video Coding (HEVC) agreed April 2010
- Document archives and software are publicly accessible
  - https://hevc-info.(general info site with links & papers, maintained by HHI)
  - http://iset.iti.darmstadt.de/ (joint effort with some links)
  - http://phenix.it.ull.unis.edu/jct
  - http://sys.tux.ch/av-arch/jct-site
- Email reflector
  - http://mailman.rwth-aachen.de/mailman/listinfo/jct-vc

Basic Technology Architecture

- All proposals were basically conceptually similar to AVC (and prior standards)
  - Block-based
  - Variable block sizes
  - Block motion compensation
  - Fractional-pel motion vectors
  - Spatial intra prediction
  - Spatial transform of residual difference
  - Integer-based transform designs
  - Arithmetic or VLC-based entropy coding
  - In-loop filtering to form final decoded picture
- Lots of variations at the individual “tool” level
- Proposal survey output documents:
  - Decoder speed JCTVC-A201
  - Architectural outline JCTVC-A202
  - Table of design elements JCTVC-A203

Technology Design Elements (part 1)

High-Level Structure

| Enhanced support for frame rate temporal sub-layer nesting and open-GOP random access |
| The structured rectangular region coding               |
| Waveform structured dependencies for prediction       |
| Enhanced reference picture set syntax                 |
| Segmentation lines and blocks                         |

Coding tree units (CTUs) are the fundamental region units (roughly analogous to macroblocks); CTU size can be 16×16, 32×32 or 64×64 luma samples (with corresponding chroma). Coding units (CUs) within CTU structure require coding unit sizes (Sxkxk) for k = 4, 8, 16, 32; Intra and Inter prediction passed at the CU level.

Prediction units: CUs for coding unit size 2Nx2N, N=4, 8, 16, 32; N=32 for 2N×32; 2N=128 for 32×128; 2N=256 for 64×2×64 and, for 2N×N, also for 2N×(N-256) and (N-256)×2N; for Intra, only 2N×2N and 4×4.

Transform units (TUs) with 4 two-dimensional 8×8 transform coding unit sizes (in 1:N ratio, for k = 4, 8, 16, 32). Intra and Inter prediction passed at the TU level.

Spatio-Temporal (SA) prediction

- Scalable Temporal Scalability
- Scalable Spatial Scalability
- Scalable Structure Scalability
- Scalable Bitstream rate scalability

Input Video Signal

General Control

Transform Scalability & Quantization

Scalable Transform Coefficients

Fiber Control Signals

Filter Control Data

Motion Data

Deblocker

Output Video Signal

Technology Design Elements (part 2)

Interpicture Prediction

- Angular intra prediction (CD decision with unified processing and prediction filtering)
- Planar surface fitting prediction

Intra-picture Prediction

- Luma motion compensation interpolation: 1.4 sample precision; 2x 3 tap separation with 7.4 tap values
- Chroma motion compensation interpolation: 1.8 sample precision, 4x 7 tap separation with 8 bit taps values
- Advanced motion vector prediction with motion vector ‘completion’ and temporal candidates
- Region merging prediction (spatial and temporal) and direct and skip modes
- No fixed prediction for 4×4, no prediction smaller than 8×8

Entropy Coding & Transformation Coefficient Coding

- CABAC coded: library of arithmetic coding CAVLC, contextual and simplified
- Transform coefficients for large transform blocks are coded in 4×4 transform coefficient groups
- Mode-dependent selection among three 4×4 scalar orders: diagonal, horizontal and vertical
- A sign bit can be “visible” in the case of each zero 4×4 transform coefficient group
- Rate-distortion optimized R-D shaping
- Deblocking filter (parallel friendly, for 8×8 edges only)
- Sample-adaptive offset filter with continuous band smoothing and directional edge refinement effects

JCT-VC Meetings & Milestones

- First “A” meeting: Dresden, Germany, 15-19 April 2010
- 586 people, 400 input documents, first full Model Under Consideration (MVC)
- Second “B” meeting: Geneva, Switzerland 21-28 July 2010
- 221 people, 300 input documents
- Third “C” meeting: Guangzhou, China, 7-13 Oct. 2010
- 244 people, 200 input documents
- Fourth “D” meeting: Daegu, Korea, 20-26 Mar. 2011
- 268 people, 200 input documents
- Fifth “E” meeting: Geneve, Switzerland, 21-27 July 2011
- 213 people, 180 input documents
- Sixth “F” meeting: Geneva, Switzerland, 21-27 Nov. 2011
- 284 people, 1000 input documents
- Eighth “H” meeting: San José, United States, 1-5 Feb. 2012
- 286 people, 200 input documents
- Ninth “I” meeting: Geneva, Switzerland, 30 April - 6 May 2012
- 246 people, 330 input documents
- Tenth “J” meeting: Stockholm, Sweden, 11-17 July 2012
- 294 people, 550 input documents (call for proposals in viability)
- Ninth “K” meeting: Shanghai, China, 4-10 Dec. 2012
- 473 people, 350 input documents (new work on scalable/layered)
- Tenth “L” meeting: Geneva, Switzerland, 16-23 Jan. 2013
- D03/HEVC AM092-a & ITU-T G1110-if0-2-draft 12
- 294 people, 450 input documents

HEVC presentation for DCC 2013

Gary J. Sullivan

2/3/2013
HEVC - High-layer syntax structure

- NAL unit structuring:
  - Similar concept as in AVC – identification of VCL payload & parameter sets
  - More NAL unit types (64 max.), currently 25 defined, 2 byte header
  - New video parameter set describing bitstream characteristics
- Enhanced support for open-GOP random access and bitstream splicing
  - Specific VCL payload types: Clean random access (CRA), broken link access (BLA), with random access decodable (RADL) and random access skipped (RASL) pictures
- Support for temporal sub-layers
  - Temporal sub-layer access (TSA) allows to identify at which points of the bitstream a change in picture rate can be made
- Reference picture set syntax
  - More explicit and robust than in AVC

HEVC - Slices, Slice Segments, Tiles, and Wavefronts

- A slice is an independently decodable entity; one picture can contain multiple slices
  - Significant number of parameters conveyed in slice header
  - A slice is often restricted by packet payload size, and therefore consists of variable number of CTUs
  - Can chop into slice segment strings of CTUs for packetization
- Tiles are also independently decodable in terms of entropy coding and intra prediction, but have a lean header, and share information from the picture level
  - Dividing a picture into regular-sized tiles (fixed number of CTUs), enables efficient parallel processing and provides entry points for local access
  - Ordered substreams for wavefront parallel processing of CTUs that are mutually independent w.r.t. CABAC adaptation

Profiles, Tiers, and Levels

- As in previous standards, HEVC will define conformance points by profile (combinations of coding tools) and levels (picture sizes, maximum bit rates etc.).
- New concept of "tiers" for bit rate and buffering capability
- A conforming bitstream must be decodable by any decoder that is conforming the given profile/tier/level combination
- 3 profiles in the first version (see next slide)
- 13 levels which cover all important picture sizes ranging from VGA at low end up to 8K x 4K at high end
  - Level 5.1 includes 4K Ultra HD @60 Hz, Level 6.1 includes 8k @60Hz
  - Most levels have two tiers: High and Main

Wavefront Parallel Processing

- Wavefront processing allows to run several processing threads in a slice over rows of CTUs with a 2-CTU delay that allows adaptation

HEVC Profiles

- "Main" profile approximately follows the overview given in the preceding slides, with following restrictions
  - Only 8-bit video with VC32 4:2:0 is supported
  - Wavefront processing can only be used when multiple tiles in a picture are not used
- "Main Still Picture" profile
  - For still-image coding applications
  - Bitstream contains only a single (intra) picture
  - Includes all (intra) coding features of Main profile
- "Main 10" profile
  - Additionally supports up to 10 bits per sample
  - Includes all coding features of Main profile
HEVC Levels and Tiers

HEVC HM 5.0 High Efficiency vs. JM 18.2 AVC High Profile (ITU-T H.264 | ISO/IEC 14496-10)

- Subjective quality is what really matters, but here are some PSNR results

- Interactive application (low delay)

- Entertainment application (random access)

- HEVC HM 5 PSNR Performance

- HEVC HM 8 PSNR Performance

- HEVC HM 8 PSNR Performance

- HEVC Levels and Tiers

- Interactive application (low delay)

- Entertainment application (random access)
HEVC HM 8 PSNR Performance

- Entertainment application (random access)

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Rate Saving HEVC HP</th>
<th>Rate Saving H264 MP</th>
<th>Rate Saving MPEG-4 ASP</th>
<th>Rate Saving H263 HLP</th>
<th>Rate Saving H264/MPEG-4 AVC HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEVC HP</td>
<td>35.4%</td>
<td>44.5%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MPEG-4 ASP</td>
<td>-</td>
<td>-</td>
<td>3.9%</td>
<td>19.7%</td>
<td>-</td>
</tr>
<tr>
<td>H263 HLP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>16.2%</td>
</tr>
<tr>
<td>H264/MPEG-4 AVC HP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>70.8%</td>
</tr>
</tbody>
</table>

Average over entire test set and all bit rates:

HEVC presentation for DCC 2013

Gary J. Sullivan 2013-03-20

HEVC HM 6 PSNR Performance: Intra only

- Image "Barbara"

<table>
<thead>
<tr>
<th>HEVC vs Other Codecs</th>
<th>PSNR Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEVC</td>
<td>15.9</td>
</tr>
<tr>
<td>JPEG2000</td>
<td>22.9</td>
</tr>
<tr>
<td>JPEG XR</td>
<td>30.4</td>
</tr>
<tr>
<td>WebP</td>
<td>31.4</td>
</tr>
<tr>
<td>JPEG</td>
<td>433</td>
</tr>
</tbody>
</table>

Source: Nguyen/Marpe (JCTVC-I0595)

HEVC presentation for DCC 2013

Gary J. Sullivan 2013-03-20

HM5 vs AVC Subjective Performance

- HEVC HM 5 vs. AVC HM18 which used similar encoder optimization, same GOP structure setting etc.

- Subjective quality is what really matters: PSNR is not able to reflect that HEVC provides better spatio-temporal consistency of the decoded video with less fluctuating artifacts - e.g. due to the larger block structures, new loop processing elements and better interpolation filters

- Emphasis in this test on delay-tolerant applications with relatively frequent random access points

- DSIS methodology was used, 9 test sequences at 4 rates each, 24 test subjects were employed in front of a full HD display at 2H viewing distance

- These test results indicate that for relevant cases a 50% or more bit rate reduction is achieved compared to AVC High Profile

HEVC presentation for DCC 2013

Gary J. Sullivan 2013-03-20
**HM5 vs AVC Subjective Performance**

- Average bit rate savings computed from interpolated "MOS over rate" graphs (from an separate test, JCTVC-H0116, Tan et al.)

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Bit rate Savings relative to H.264/AVC HP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HEVC MF</td>
</tr>
<tr>
<td>BQTerrace</td>
<td>65.1%</td>
</tr>
<tr>
<td>BasketballDrive</td>
<td>66.6%</td>
</tr>
<tr>
<td>Kimono1</td>
<td>55.2%</td>
</tr>
<tr>
<td>ParkScene</td>
<td>49.7%</td>
</tr>
<tr>
<td>Cactus</td>
<td>50.2%</td>
</tr>
<tr>
<td>BQMall</td>
<td>41.6%</td>
</tr>
<tr>
<td>BasketballDrib</td>
<td>44.9%</td>
</tr>
<tr>
<td>PartyScene</td>
<td>29.8%</td>
</tr>
<tr>
<td>RaceHorses</td>
<td>42.7%</td>
</tr>
<tr>
<td>Average</td>
<td>49.3%</td>
</tr>
</tbody>
</table>

**HEVC HM 5 Subjective Performance for Low-Delay Applications**

- **Latest HEVC HM 5.0 High Efficiency vs. JM 18.2 AVC High Profile**
  (ITU-T H.264 / ISO/IEC 14496-10)
- **Source:** JCTVC-H0362 by F. Kossentini, N. Mahdi, H. Guermazi, M. Horowitz, S. Xu, B. Li, G. J. Sullivan, J. Xu (Jan. '12)
- **Subjective quality** should be measured formally with proper statistical analysis and controlled viewing, but here are some **informal** test results
- Some encoding methods as described in JCTVC-H0360 (similar to prior JCTVC-0399)
- Five video test sequences (three Class E 720p, two Class B 1080p)
- Encodings and bit rates selected to represent low-delay applications
- Compared AVC encodings with HEVC encodings at half the bit rate
- In 75% of test cases, viewers either had no preference or preferred HEVC
- For 4 of 5 video clips, most viewers had no preference or preferred HEVC
- For 3 of 5 videos, the vast majority who had a preference preferred HEVC (and for one other clip, the preferences rate was nearly equal)
- These test results are not officially endorsed by JCTVC or others

**Low-Delay Subjective Assessment Results:** SPIE 8499-31 (Horowitz et al)

- **Reference encoders:** HM 7.1 vs. JM 18.3

<table>
<thead>
<tr>
<th>Subjective viewing for JM vs. HM</th>
<th>Subjective viewing for eBrisk vs. x264</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence</td>
<td>HEVC bit rate (Mbits)</td>
</tr>
<tr>
<td>Clarice</td>
<td>149 36 382 37</td>
</tr>
<tr>
<td>Video1</td>
<td>190 37 367 36</td>
</tr>
<tr>
<td>OliveTreeCreek</td>
<td>408 37 870 37</td>
</tr>
<tr>
<td>Kimono1</td>
<td>462 36 1484 35</td>
</tr>
<tr>
<td>race_and_calendar</td>
<td>347 37 754 38</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>489 36 92 33</td>
</tr>
</tbody>
</table>
**Complexity & Deployment**

- Overall complexity report SPIE 8499:32 (Ahn, Han, & Sim)
  - Based on reference software analysis
  - 25-30% decoder complexity increase relative to H.264/AVC HP
  - Parallelism additionally improved
- Decoder demos by F. Bossen of NTT DoCoMo in July & October (JCTVC-J0128 & JVTVC-K0327 & JCTVC-L0098)
  - 4k x 2k at 30 fps on a laptop single-threaded
  - 1920 x 1080p at 25 fps on a smart phone single-threaded
  - 1280 x 720 at 30 fps on ARM Cortex A9 clocked at 1 GHz (iPad)
  - 4k x 2k at 60 fps (up to 100 fps) on a laptop three-threaded
- Encoding complexity is more of a challenge
  - But feasible (Eltrok, Ericsson, Vanguard, Allegro, Rovi, Ateme, NGCodec, Elemental, etc.)
  - Design structure and principles similar to H.264/AVC
  - Design is flexible for selection of aspects to support
  - Parallelism opportunities built in

**Summary and outlook**

- Very active project (500+ documents, 750+ people)
- Very diverse company & university participation (~150 institutions)
- Major technical advance over prior standards
- Computational/implementation complexity is reasonable
  - Parallelism is an increased theme
- Three profiles in first version, with two bit rate tiers and 13 levels
- Deliverables:
  - Video coding specification
  - Reference software
  - Conformance testing specification
  - Systems support under way for MPEG-2 TS, ISO BMFF, DASH, etc.
  - Patent pool formation begun in MPEG-LA
  - Multiple versions and extensions planned (Rext, 3D/MVC, SVC)
- Contact: JVT, JCT-VC, JCT-3V, VCEG, MPEG video chairs:
  - Gary J. Sullivan (garysull@microsoft.com)
  - Jens-Rainer Ohm (ohm@ient.rwth-aachen.de)

**HEVC extension developments**

- Range extensions (JCTVC-L1005 draft)
  - Increased bit depths
- Two core experiments underway (JCTVC-L1121 and JCTVC-L1222)
- Scalability “SHVC” (JCTVC-L1008 draft)
  - Hooks for extensions built into version 1
  - Joint call for proposals in 2012
  - 21 proposals received
  - Spatial & SNR enhancements planned
  - Multi-loop coding structure likely
  - AVC base layer possible
  - Five core experiments underway (JCTVC-L1101 to JCTVC-L1105)
- 3D (New JCT-3V partnership)
  - Frame packing in version 1
  - MPEG call for proposals in 2011
  - Multiview & depth map encoding & combined encoding
  - Extensions to AVC as well as HEVC (first extension finished, more)
  - Third Multiview HEVC draft produced (JCTVC-X1003)
  - Seven core experiments underway (JCTVC-C1001 to JCTVC-C1107)

**For further information**

- Document archives and software are publicly accessible
  - [http://hevc.info](http://hevc.info) (general info site with links & papers, maintained by HHI)
  - [http://jct-vc.org](http://jct-vc.org) (http://www itu.int/ITU-T/videoprocessing/changes/jct-jctvc.aspx)
  - [http://hevc-info.sudparis.eu/jct](http://hevc-info.sudparis.eu/jct)
  - [http://hevc-info.sudparis.eu/jct3v](http://hevc-info.sudparis.eu/jct3v)
  - [http://ftp.itu.int/arch/jctvc/site](http://ftp.itu.int/arch/jctvc/site)

- Publications
  - Special Issue on Emerging Research and Standards in Next Generation Video Coding (HEVC), IEEE T-CSVT, Dec. 2012 (includes technical overview paper, compression capability analysis paper, complexity analysis paper, etc)
  - Special Section on the Joint Call for Proposals on High Efficiency Video Coding (HEVC) Standardization, IEEE T-CSVT, Dec. 2010
  - “Recent Developments in standardization of High Efficiency Video Coding (HEVC), SPIE Appl. Dig. Image Proc., Aug. 2010"