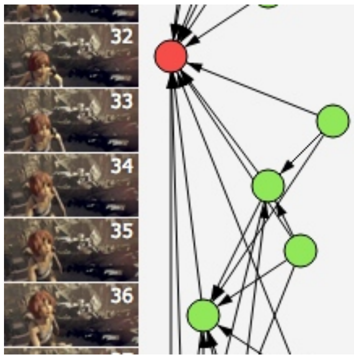
HEVC:
An introduction to High Efficiency Video Coding



About Vcodex

Vcodex are world experts in video compression. We provide essential analysis and advice on technology, strategy and intellectual property. Our input will help you get the most out of your video compression technology.

Video compression is the technology behind moving digital images. It is essential to video on phones, cameras, laptops and TV. In fact, anything you can watch on a screen uses video compression.

1 Summary

High Efficiency Video Coding (HEVC) is a new standard for video compression that has the potential to deliver better performance than earlier standards such as H.264/AVC.

Source video, consisting of a sequence of video frames, is encoded or compressed by an HEVC video encoder to create a compressed video bitstream. The compressed bitstream is stored or transmitted. A video decoder decompresses the bitstream to create a sequence of decoded frames.

HEVC has the same basic structure as previous standards such as MPEG-2 Video and H.264/AVC. However, HEVC contains many incremental improvements such as:

- More flexible partitioning, from large to small partition sizes
- Greater flexibility in prediction modes and transform block sizes
- More sophisticated interpolation and deblocking filters
- More sophisticated prediction and signalling of modes and motion vectors
- Features to support efficient parallel processing.

The result is a video coding standard that can enable better compression, at the cost of potentially increased processing power.

2 What is HEVC?

1. An international standard for video compression. Developed by a working group of ISO/IEC MPEG (Moving Picture Experts Group) and ITU-T VCEG (Video Coding Experts Group), HEVC is an international standard, jointly published as ISO/IEC 23008-2 and ITU-T Recommendation H.265. HEVC is published as a document (the standard itself) together with a reference software implementation (the test model, HM).

2. A format for compressed video. The HEVC standard specifies a format for compressed or encoded video sequences, together with a method for decoding this format. An HEVC-compatible video sequence should (a) meet the specification of the compressed video format and (b) be correctly decodeable using the method described in the standard. HEVC video sequences can be stored in media files, streamed over the internet, transmitted by broadcast, etc.

3. A set of tools or methods for video compression. HEVC specifies a number of methods or tools that may be used by a video compression encoder. It’s up to the designer of the encoder which tools are actually used, and how they are applied

4. Better video compression. Depending on how the tools are used, HEVC has the potential to offer significantly higher compression than earlier standards such as H.264 / AVC. Achieving the best possible compression is likely to require significant computational resources.

3 Why do we need it?

HEVC aims to provide a step change improvement in video compression compared with earlier standards. HEVC's predecessor, the H.264/AVC standard, was first published in 2003. Since then, digital video has become increasingly ubiquitous. High Definition is now the norm for many devices and applications. HEVC was developed to address the following trends:

- Widespread use of digital video, at increasingly high resolutions, which puts a significant strain on network capacity.
- Increasing use of video resolutions beyond HD, which will increase the burden on networks and storage even further.
- Continuing improvements in processing capacity. In 2013, a mobile handset or tablet is likely to have more computing power than a desktop computer from 2003.

With these issues in mind, a new video compression standard that makes use of higher computational capacities to enable more efficient handling of high resolution video is an attractive proposition. With HEVC, it should be possible to store or transmit video more efficiently than with earlier technologies such as H.264. This means:

- At the same picture size and quality, an HEVC video sequence should occupy less storage or transmission capacity than the equivalent H.264 video sequence.
- At the same storage or transmission bandwidth, the quality and/or resolution of an HEVC video sequence should be higher than the corresponding H.264 video sequence.

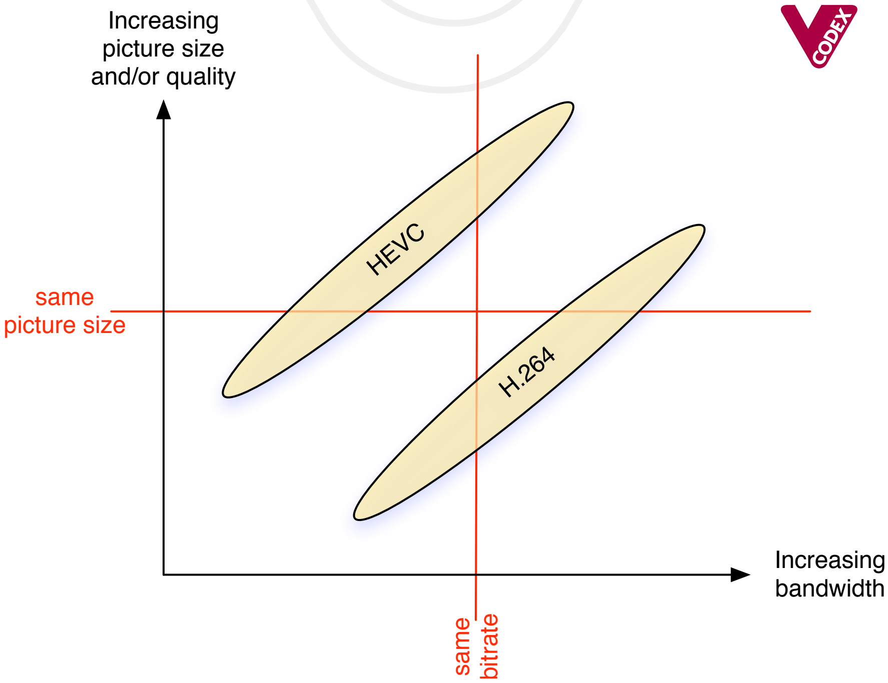


Figure 1: The potential gains of HEVC vs. H.264 (not to scale)

4 How does HEVC work?

HEVC is based on the same general structure as previous standards. Source video, consisting of a sequence of video frames, is encoded or compressed by a video encoder to create a compressed video bitstream. The compressed bitstream is stored or transmitted. A video decoder decompresses the bitstream to create a sequence of decoded frames.

The steps carried out by a video encoder (Figure 2) include:

- Partitioning each picture into multiple units
- Predicting each unit using inter or intra prediction, and subtracting the prediction from the unit
- Transforming and quantizing the residual (the difference between the original picture unit and the prediction)
- Entropy encoding the transform output, prediction information, mode information and headers.

A video decoder reverses the steps:

- Entropy decoding and extracting the elements of the coded sequence
- Rescaling and inverting the transform stage
- Predicting each unit and adding the prediction to the output of the inverse transform
- Reconstructing a decoded video image.

The HEVC standard defines (ii) the syntax or format of a compressed video sequence and (ii) a method of decoding a compressed sequence. The actual design of the encoder is not standardised.

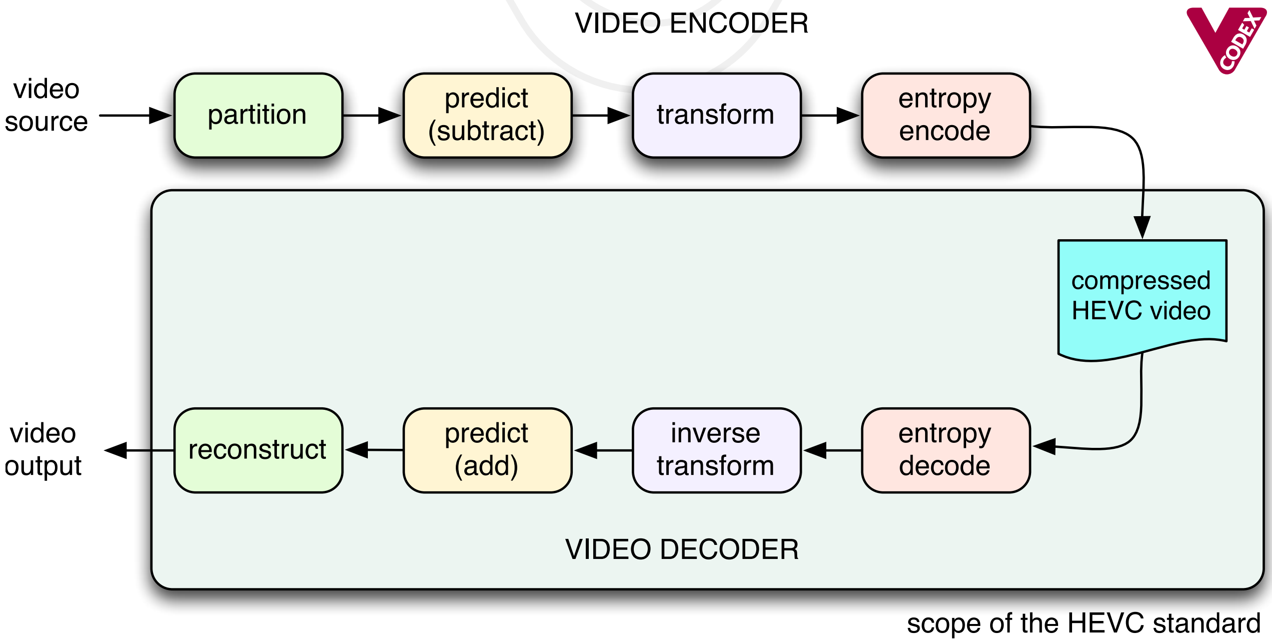


Figure 2: Structure of an HEVC encoder and decoder

4.1 Partitioning

HEVC supports highly flexible partitioning of a video sequence. Each frame of the sequence is split up into rectangular or square regions (Units or Blocks), each of which is predicted from previously coded data. After prediction, any residual information is transformed and entropy encoded.

Each coded video frame, or picture, is partitioned into Tiles and/or Slices, which are further partitioned into Coding Tree Units (CTUs). The CTU is the basic unit of coding, analogous to the Macroblock in earlier standards, and can be up to 64x64 pixels in size.

A Coding Tree Unit can be subdivided into square regions known as Coding Units (CUs) using a quadtree structure (Figure 3). Each CU is predicted using Inter or Intra prediction and transformed using one or more Transform Units (see below).

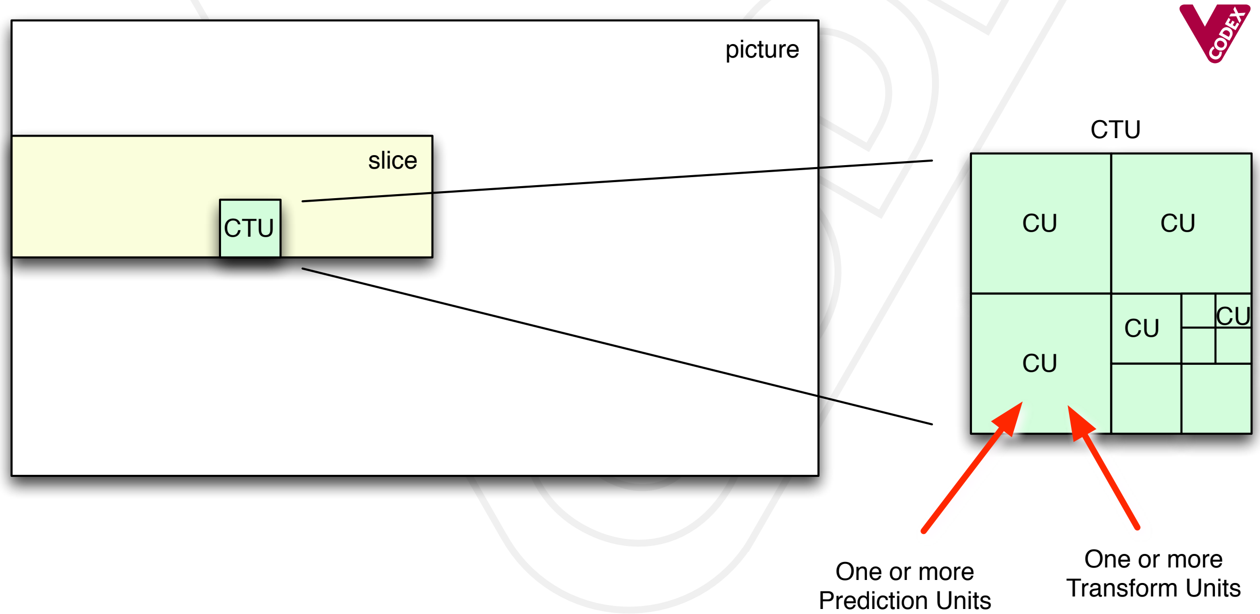


Figure 3: Picture, slice, Coding Tree Unit (CTU), Coding Units (CUs)

Figure 4 shows a video frame partitioned into slices, with one slice highlighted in blue. The highlighted slice contains six 64x64 CTUs.

Figure 4: Video frame showing Slices and Coding Tree Units (source: Parabola Research)

Figure 5 shows a close-up of the CTU highlighted in Figure 4. The 64x64 CTU is split into four 32x32 regions, with the top-left 32x32 CU highlighted. In the other four quarters, the 32x32 region is split further, to 16x16 or 8x8 CUs.

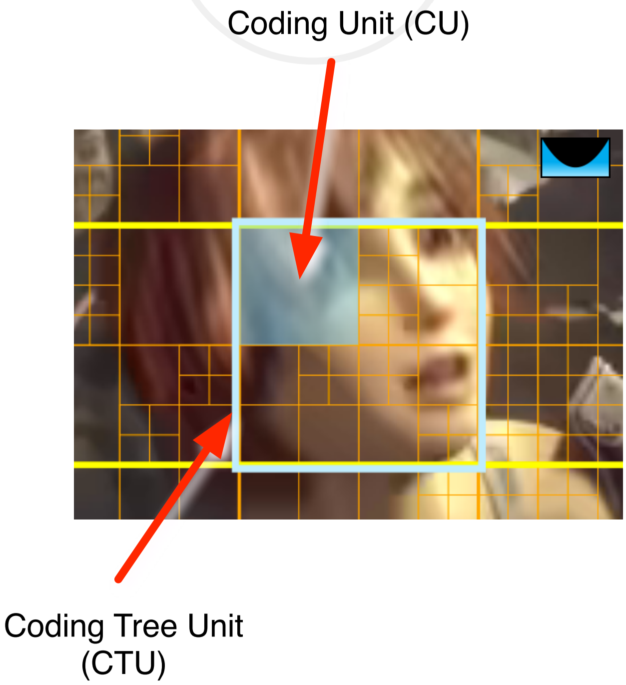


Figure 5: Coding Tree Unit subdivided into Coding Units (source: Parabola Research)

4.2 Prediction

Frames of video are coded using Intra or Inter prediction. Figure 6 shows a sequence of coded video frames or coded pictures. The first picture (0) is coded using Intra prediction only, using spatial prediction from other regions of the same picture. Subsequent pictures are predicted from one, two or more reference pictures, using Inter and/or Intra prediction for each Prediction Unit (PU). The prediction sources for each picture are indicated by arrows.

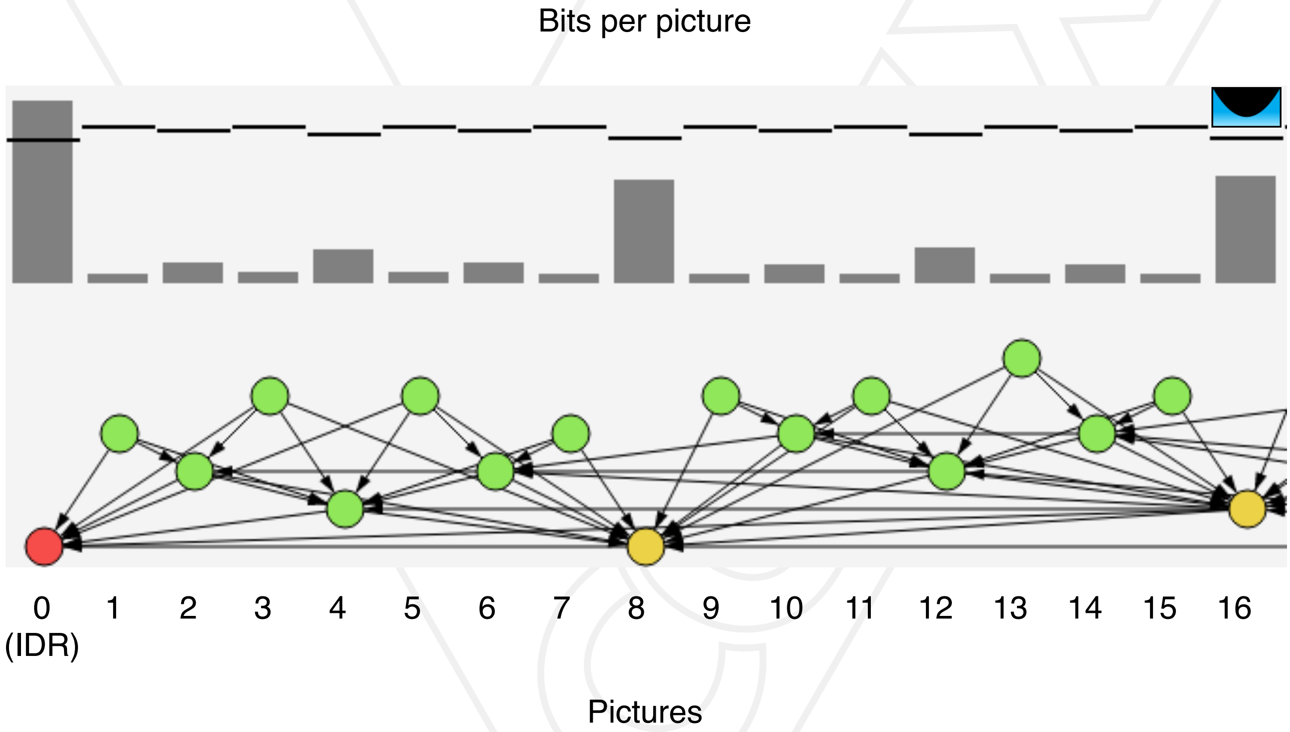


Figure 6: Sequence of coded pictures (source: Parabola Research)

Each Coding Unit (CU) is partitioned into one or more Prediction Units (PUs), each of which is predicted using Intra or Inter prediction.

Intra prediction: Each PU is predicted from neighbouring image data in the same picture, using DC prediction (an average value for the PU), planar prediction (fitting a plane surface to the PU) or directional prediction (extrapolating from neighbouring data).

Inter prediction: Each PU is predicted from image data in one or two reference pictures (before or after the current picture in display order), using motion compensated prediction. Motion vectors have up to quarter-sample resolution (luma component).

Figure 7 shows two examples of Prediction Units. The CTU in the centre of the Figure is predicted using a single 64x64 PU. All the samples in this PU are predicted using the same motion compensated inter prediction from one or two reference frames. Shown on the right is an 8x16 PU, which is part of the prediction structure for a 32x32 CU.

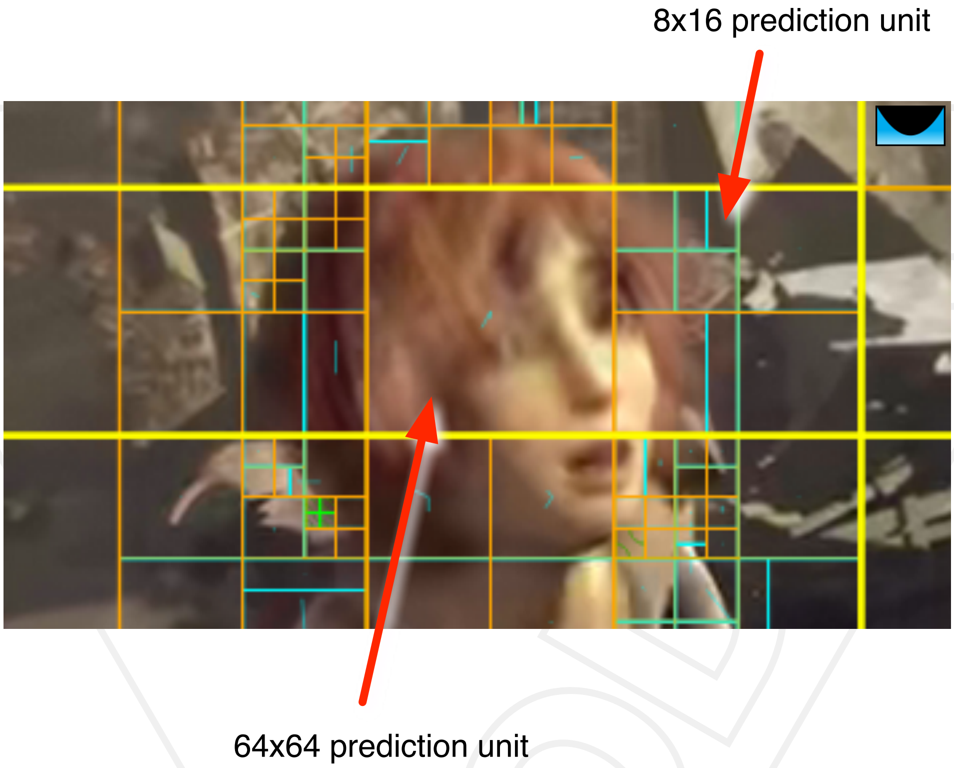


Figure 7: Two examples of inter Prediction Units (source: Parabola Research)

4.3 Transform and quantization

Any residual data remaining after prediction, is transformed using a block transform based on the Discrete Cosine Transform (DCT) or Discrete Sine Transform (DST). One or more block transforms of size 32x32, 16x16, 8x8 and 4x4 are applied to residual data in each CU. Figure 8 shows the CUs in a CTU and the transforms applied to each CU. The size of each transform is indicated by the size of the circles.

The highlighted 8x8 CU is processed with an 8x8 block transform and quantized. After quantization, any remaining non-zero transform coefficients are scanned in a zigzag order. In this case, only four non-zero coefficients remain after prediction, transform and quantization. Other CUs in this CTU are processed with 32x32, 16x16, 8x8 or 4x4 transforms, indicated by the size of each circle. In the case of the lower-right CU, no residual data remains after prediction, transform and quantization, so no transform coefficients are encoded.

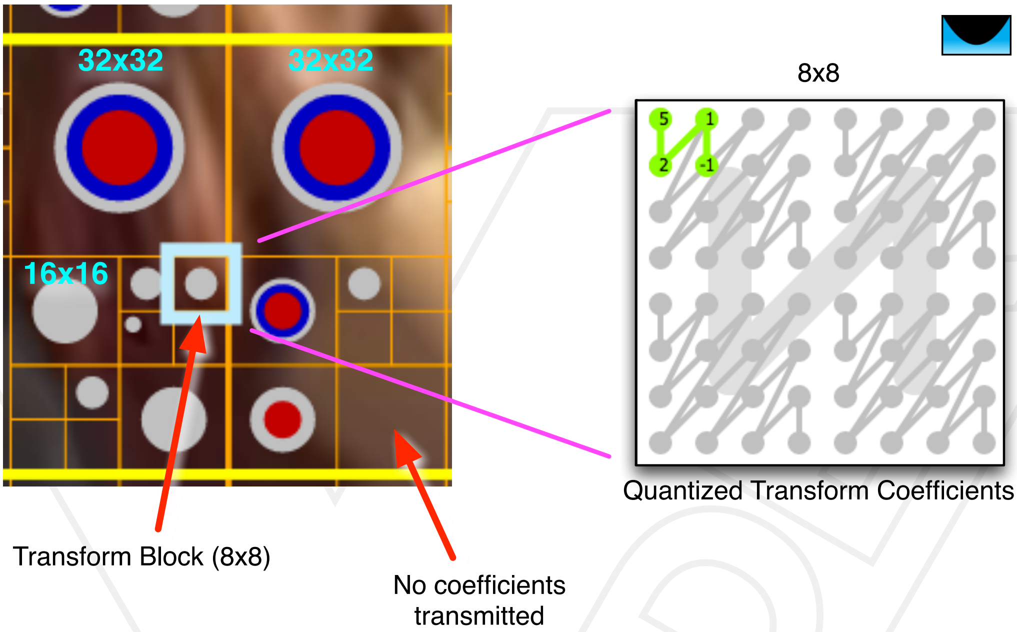


Figure 8: CTU showing a range of transform (TU) sizes (source: Parabola Research)

4.4 Entropy coding

A coded HEVC bitstream consists of quantized transform coefficients, prediction information such as prediction modes and motion vectors, partitioning information and other header data. All of these elements are encoded using Context Adaptive Binary Arithmetic Coding (CABAC).

Figure 9 shows an Inter-coded video frame with an overlay representing the number of coded bits per Coding Unit. In this example, the partitions tend to follow the underlying structure of the video scene. Smaller CUs tend to be used around complex edges and moving objects, such as the figure of the girl. In general, the encoder generates more coded bits around moving and changing parts of the scene.

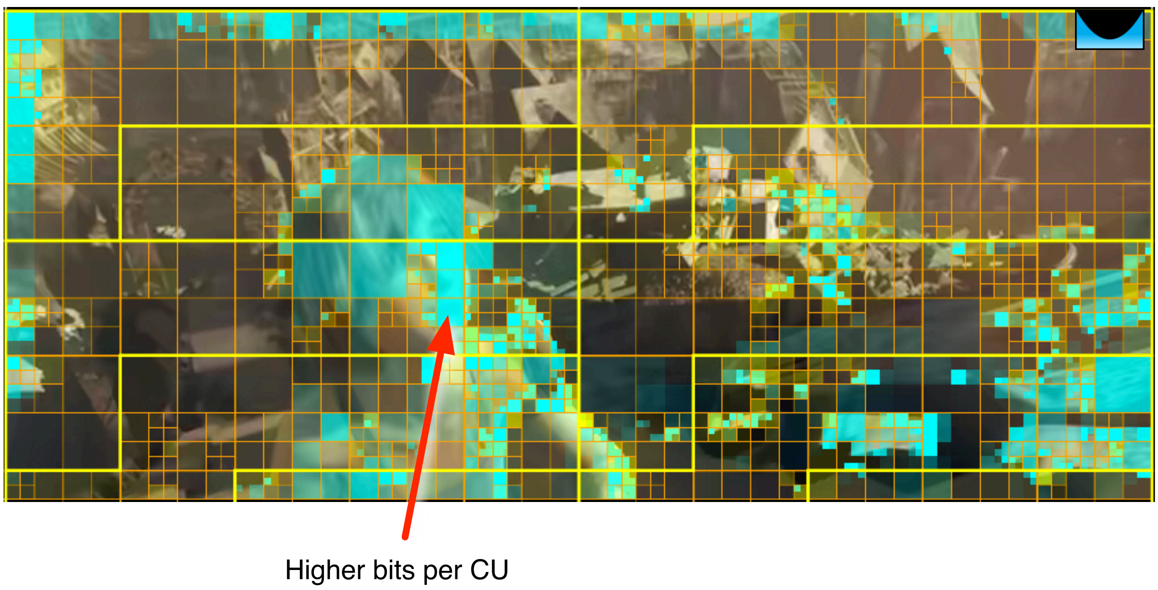


Figure 9: Video frame showing bits per CU (source: Parabola Research)

4.5 Other features

Mode and motion vector prediction: HEVC features sophisticated prediction and merging of mode information, based on the modes of previously-coded units.

Deblocking filter: A filter is applied to luma and chroma samples next to TU or PU boundaries (where these boundaries are aligned on an 8x8 grid). The strength of this filter may be controlled by syntax elements signalled in the HEVC bitstream. The deblocking filter is intended to reduce visual artifacts around block / unit edges that may be introduced by the lossy encoding process.

Sample Adaptive Offset: An optional filter that enables adjustment of the decoded video frames and can enhance the appearance of smooth regions and edges of objects. The SAO filter is a non-linear filter that makes use of look-up tables that may be signalled in the HEVC bitstream.

Parallel processing: HEVC includes a number of features that may be useful for decoders with parallel processing capabilities. Tiles are rectangular regions of a picture that may be decoded largely independently. Wavefront Parallel Processing (WPP) is an encoder mode that ensures that a new row of CTUs can be decoded after only two CTUs have been decoded in the previous row. The corresponding syntax elements may be mapped to separate Network Abstraction Layer (NAL) units, and hence separate network packets.

Profiles, Levels and Tiers: A Profile determines a subset of the available HEVC coding tools that must be supported by a decoder. The combination of Level and Tier specifies maximum decoder processing capabilities in terms of picture size, coded samples per second, bit rate and buffering.

4.6 Terminology

H.264 terminology	HEVC terminology	What it means
Frame	Frame	A complete video frame
Macroblock (MB)	CTU	Basic coding unit, a square region
Block	Coding Unit (CU)	A subset of a MB/CTU
MB partition	Prediction Unit (PU) or Prediction Block (PB)	A rectangular area predicted using intra or inter prediction
Block (transform)	Transform Unit (TU) or Transform Block (TB)	A block of samples to be transformed
Slice	Slice	A (usually) continuous sequence of MBs/CTUs
-	Tile	A rectangular set of CTUs that can be decoded in parallel

5 How good is HEVC?

Compared with previous standards such as MPEG-2 Video and H.264/AVC, HEVC can enable better compression, potentially at the cost of increased processing power.

Figure 10 shows close-ups of two decoded video frames. The same sequence (Kristen and Sara, 720p resolution) was encoded using H.264 High Profile (left) and HEVC (right) at approximately the same bitrate (420kbps). The quality of the HEVC clip is clearly better : for example, the H.264 closeup loses much of the detail of the hair and has obvious distortions in the face area.

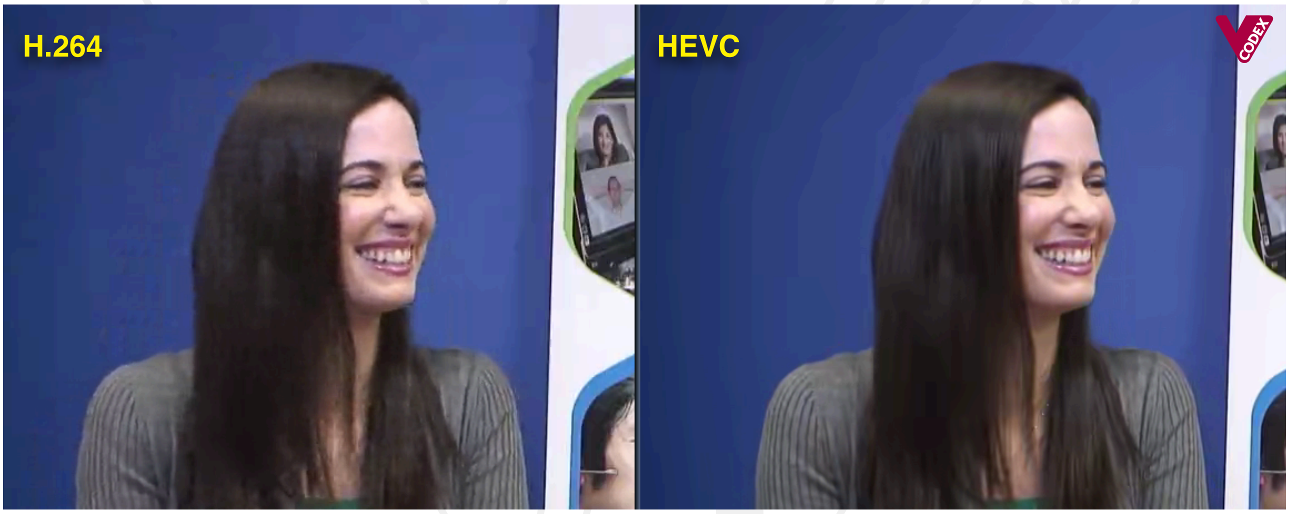


Figure 10: Close-up of sample frame encoded at 420kbps using H.264 (left) and HEVC (right)

Just how much better is HEVC than earlier standards? This depends very much on the characteristics of the video clip, on the design of the video encoder and on the opinion of the viewer. Several studies have concluded that HEVC can deliver similar quality to H.264 at approximately half the bitrate (see the references below).

6 To find out more

The draft standard: High Efficiency Video Coding Draft 10, Document JCTVC-L1003, available at:
http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=7243

HEVC reference software (HM) and software manual:
http://hevc.hhi.fraunhofer.de/

Overview of HEVC:
G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Trans. Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649-1668, Dec. 2012.

Quality evaluation:
http://www.slideshare.net/touradj_ebrahimi/subjective-quality-evaluation-of-the-upcoming-hevc-video-compression-standard

Test bitstreams:
ftp://ftp.kw.bbc.co.uk/hevc/hm-10.0-anchors/bitstreams ("Anchor bistreams")

GPAC software player, with instructions for getting started:
http://vcodex.blogspot.co.uk/2013/04/comparing-hevc-and-h264-quality-using.html

HEVC analysis software:
http://www.parabolaresearch.com/ Parabola Explorer

Acknowledgements

Thanks to Parabola Research for permission to use screenshots of 'Parabola Explorer'.

About the author

Vcodex Ltd is led by Professor Iain Richardson, an internationally known expert on the MPEG and H.264 video compression standards. Based in Aberdeen, Scotland, he frequently travels to the US and Europe.

Professor Richardson is the author of *The H.264 Advanced Video Compression Standard*, a widely cited work in the research literature. He has written three further books and over 80 journal and conference papers on image and video compression. He regularly advises companies on video codec technology, video coding patents and mergers/acquisitions in the video coding industry.

Iain Richardson
iain@vcodex.com
Vcodex.com