Image/Video Processing and Coding

Dr.-Ing. Henryk Richter
Institute of Communications Engineering
Phone: +49 381 498 7303, Room: W 8220
EMail: henryk.richter@uni-rostock.de

Content

- Introduction
- Colors, Image Matrix, Image Processing and Recognition structure
- Image Sampling and Quantization
- Discrete Image Characteristics
- Image Transformation
- Image Improvement
- Image Segmentation
- Features, Extraction, Descriptors
- Pattern Recognition (Basics, Systems for classification, Neural Networks)
- Data compression fundamentals
- Methods, techniques and algorithms for data compression
  - Data reduction, Coding, Decorrelation
- Image and Video coding standards and their specifics
  - JPEG, JPEG-2000
  - Video Coding: H.265

Introduction

- joint project JVT of the ITU-T VCEQ and MPEG
- nomenclature
  - ITU-T Rec. H.265
  - ISO/IEC 23008-2 HEVC
- development based on H.264/AVC
- specifically targeted at UHDTV video resolution (4k,8k)
  - >8 Bit per pixel included in initial specification (8 Bit or 10 Bit, RXext provides 14 Bit per component)
  - YCbCr subsampling 4:2:0 (in addition, 4:0:0, 4:2:2 and 4:4:4 in RXext)

⇨ Development goal: compression efficiency doubled (again) in comparison with previous standard

Key Changes in comparison to H.264

- Coding Tree Units (CTU) as replacement for the traditional 16x16 Macroblock
  - 16x16,32x32 or 64x64 luma samples per CTU
- Prediction Units (PU)
  - adaptive subdivision of CTU sized blocks as needed
- Transform Units (TU)
  - adaptive subdivision of CTU sized blocks as needed
  - 4x4,8x8,16x16 and 32x32 transform sizes (quantized DCT), 4x4 quantized DST
  - transform block sizes decoupled from PU sizes
- Advanced Motion Vector prediction
  - more MV prediction candidates compared to earlier standards
  - merge mode for MV coding
  - improved „skipped“ and „direct“ motion inference
- single-stage Motion Compensation
  - 7-tap or 8-tap filters for interpolation (instead of 6-tap plus bilinear interpolation)
- Intra Prediction with 35 different modes (33 directional, DC, Plane)
- Sample-Adaptive Offset (SAO)
  - non-linear amplitude mapping by lookup-table after the deblocking filter, parameters from bitstream
- Parallel Encoding Decoding support
  - Tiles, Wavefront Parallel Processing (including Entropy decoding), Dependent Slices
- Clean Random Access (CRA)
  - Open GOP principle, start with a temporal independent picture (RAP) and discard non-decodable pictures

Profiles / Levels

- initially only the Main Profile was specified for the first version of HEVC
  - acknowledgement that traditionally separate services (broadcast, mobile, streaming) converge toward multipurpose receiver devices
  - restrictions
    - only 8 Bit video with 4:2:0 chroma sampling
    - usage of tiles excludes wavefront parallel processing
    - tiles must be at least 256 x 64 luma samples large

- 13 Levels for initial specification
  - Level 1-3: SDTV resolution and below
  - Level 3.1: up to 720p @ 33 FPS
  - Level 4,4.1: HDTV 1080i/p @ 30,60 FPS
  - Level 5,5.1,5.2: 4k @ 30,60,120 FPS
  - Level 6,6.1,6.2: 8k @ 32,64,128 FPS
  - max. 6 pictures in decoded picture buffer (DPB) at maximum pixel count allowed in level, total limit of 16 pictures in DPB

High Level Syntax

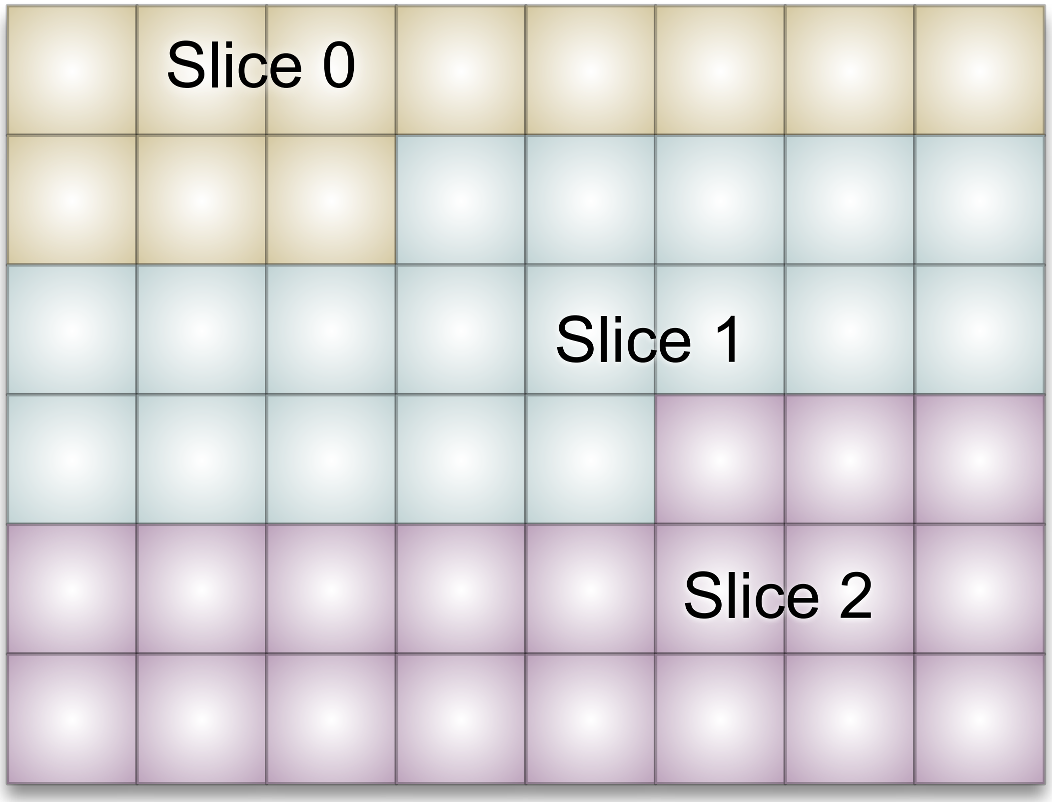
- Concepts from H.264 retained
- Network Adaptation Layer (NAL)
  - encapsulation of VCL (video coding layer) units into various transport layers (RTP, ISO MP4, MPEG-2 Systems)
- NAL units classified into VCL and non-VCL
  - VCL NAL types
    - used for different picture categories (especially random access markers)
    - temporal scalability (temporal sub-layers)
  - non-VCL NAL types
    - parameter sets
    - sequence/bitstream delimiting
    - SEI messages
      - supplemental enhancement information for parameters not directly associated with bitstream decoding
      - aspect ratio, cropping, interlace support

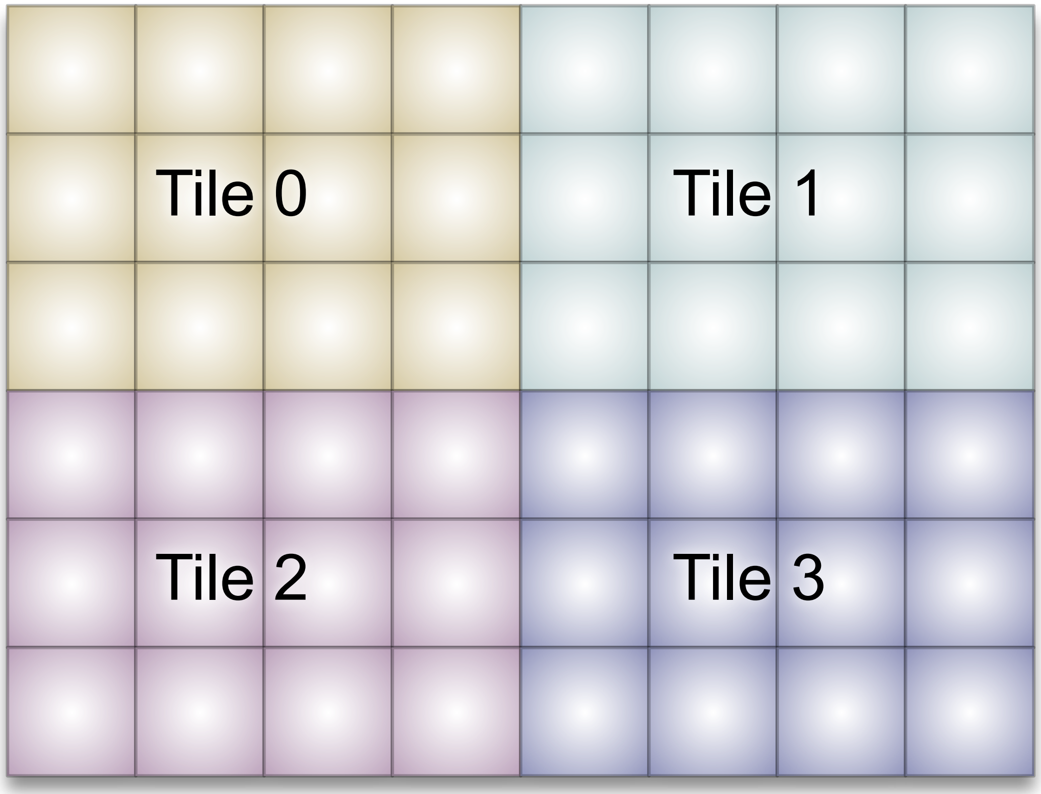
Picture Random Access

- Traditional video coders required to start decoding with I-frames (or IDR in H.264)
  - Closed GOP concept, where each GOP starts with I/IDR pictures and implicitly invalidates the reference picture buffer immediately
- HEVC Decoders can start at different random access point (RAP) pictures
  - IDR = Instantaneous Decoder Refresh
  - CRA = Clean Random Access
  - VLA = Broken Link Access
- Clean Random Access (CRA) syntax
  - Intra pictures (CRA) at the location of a random access point (RAP) to start successful decoding without the knowledge of prior pictures in the bitstream
  - Pictures following CRAs in decoding order and precede them in display order are discarded by decoders starting with CRA pictures (TFD=tagged for discard NAL unit types)
- Broken Link Access (BLA)
  - bitstream splicing, i.e. switch from one bitstream to another
  - BLA NAL units signal a disrupted picture numbering to the decoder

Frame subdivision concepts

- Slices
  - Slices consist of a sequence of CTUs
  - Arbitrary number of CTUs per slice
  - Slices can be decoded independently*
- Tiles
  - Wimplified version of the H.264 FMO framework
  - Rectangular groups of CTUs (typically same size)
  - Independent decoding
  - Each tile may contain a variable number of slices (also possible are multiple tiles within a single slice)
- Dependent Slices
  - Slices that depend on a particular decoding order of other slices (e.g. Wavefront entry points for parallel processing)

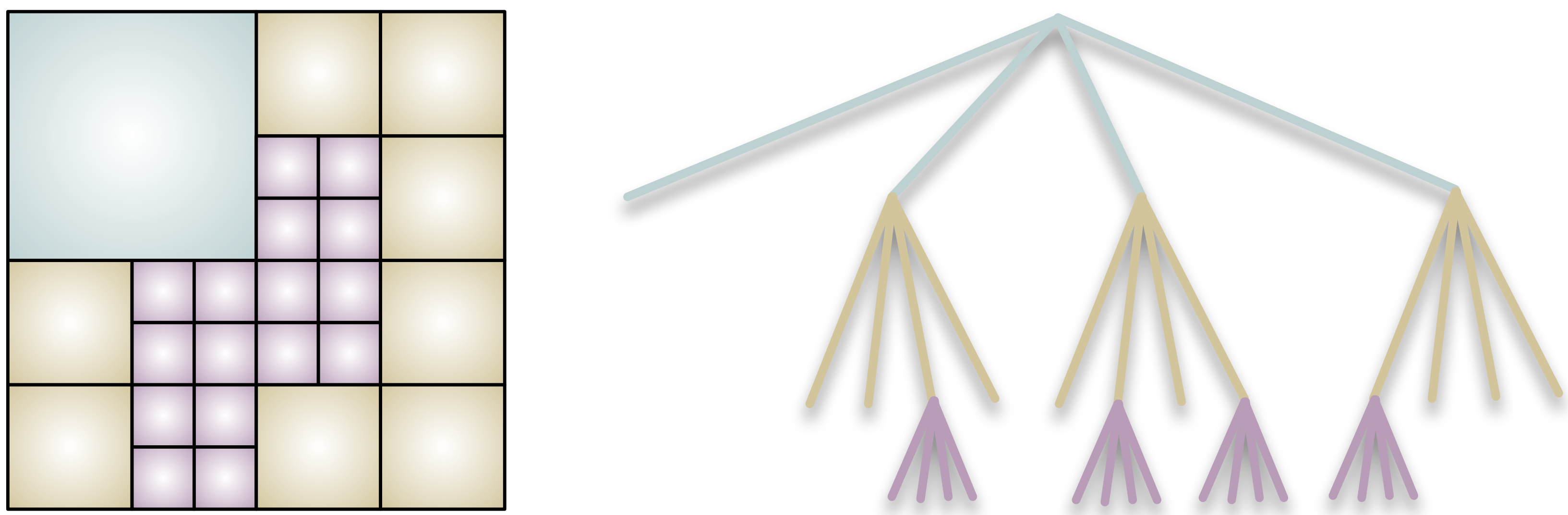




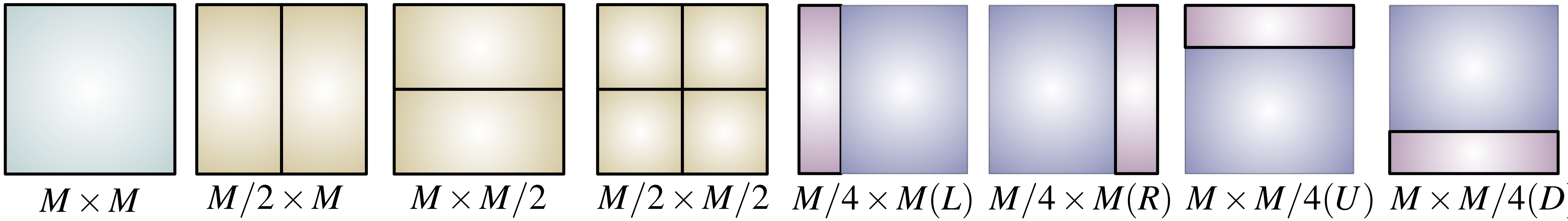
* except deblocking filter

Coding Blocks and Prediction Blocks from Coding Tree Units

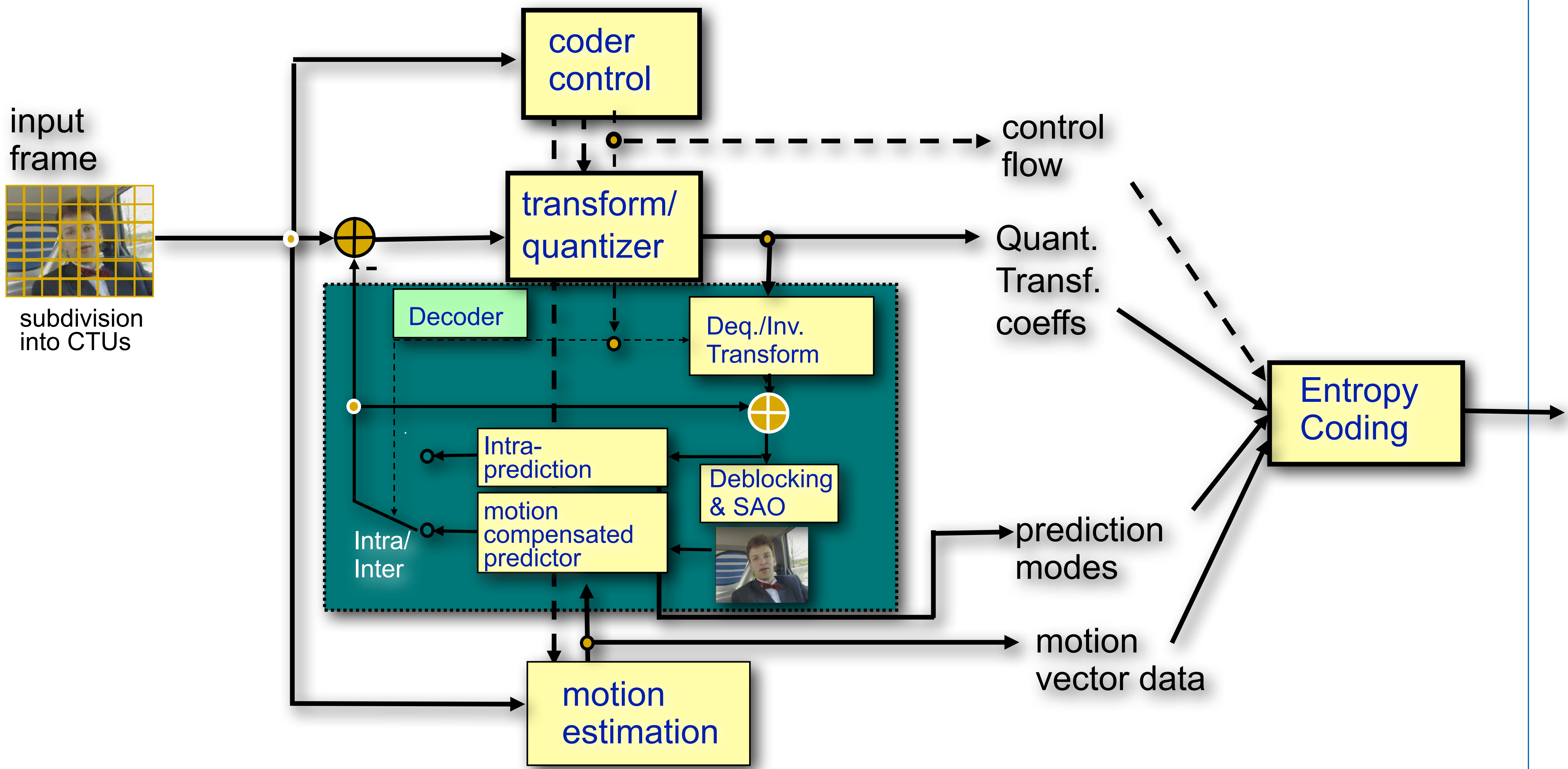
- CTUs >16x16 are the key feature of H.265 towards higher coding efficiency
- Quadtree decomposition of CTUs into smaller Coding Blocks (CB)



- Recursive quadtree partitioning of CBs into Transform Blocks (TB) (min. size 4x4)
- In Intra CUs, each CB may be (optionally) further subdivided into 4 quadrants, where each quadrant is assigned a distinct intra prediction mode (min. size 4x4)
- In Inter CUs, the CB may be (optionally) subdivided into two prediction blocks (PB), or into four PBs when the CB size is at the minimum allowed CB size



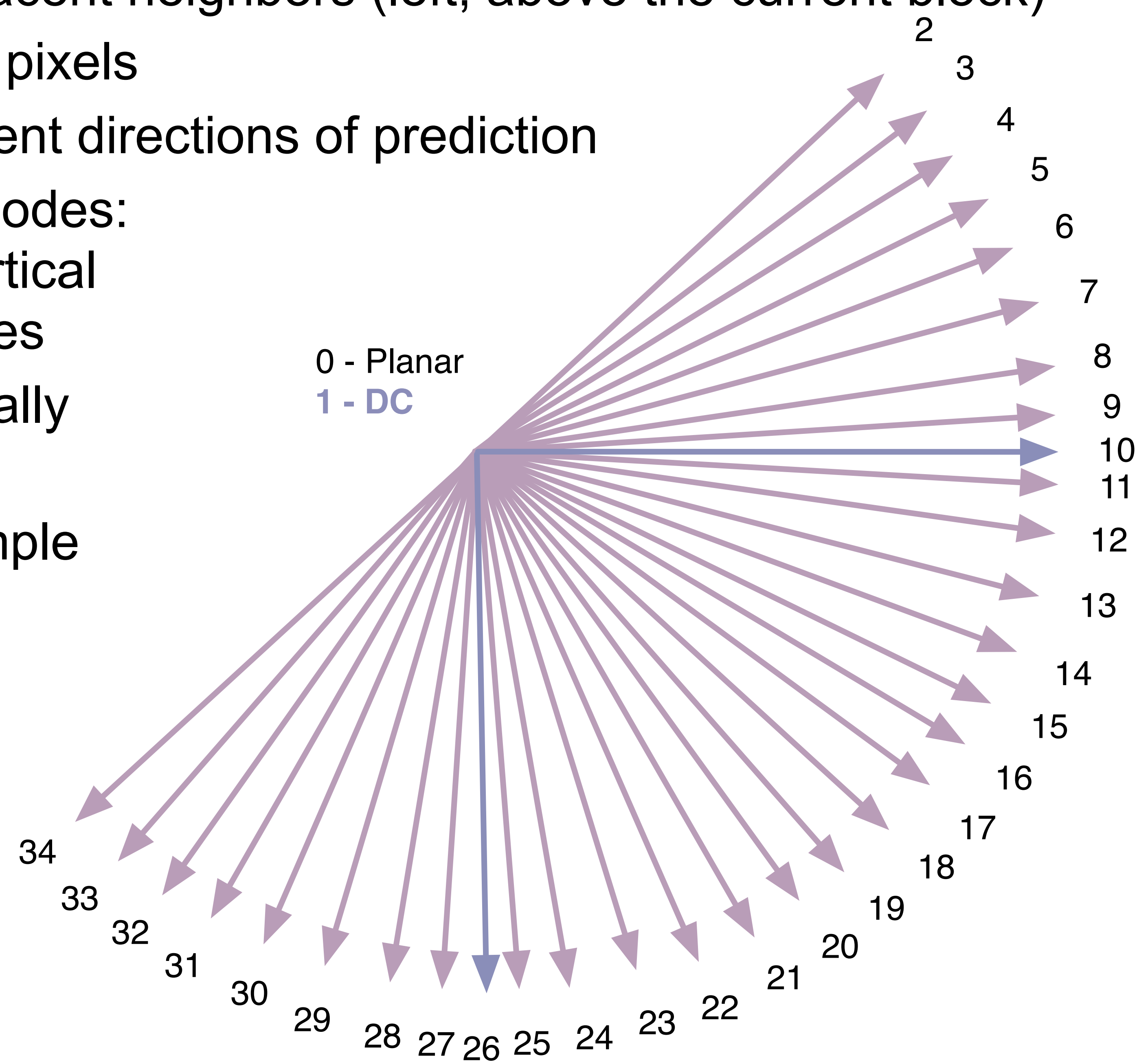
H.265 / HEVC system architecture



Quelle: HHI Berlin

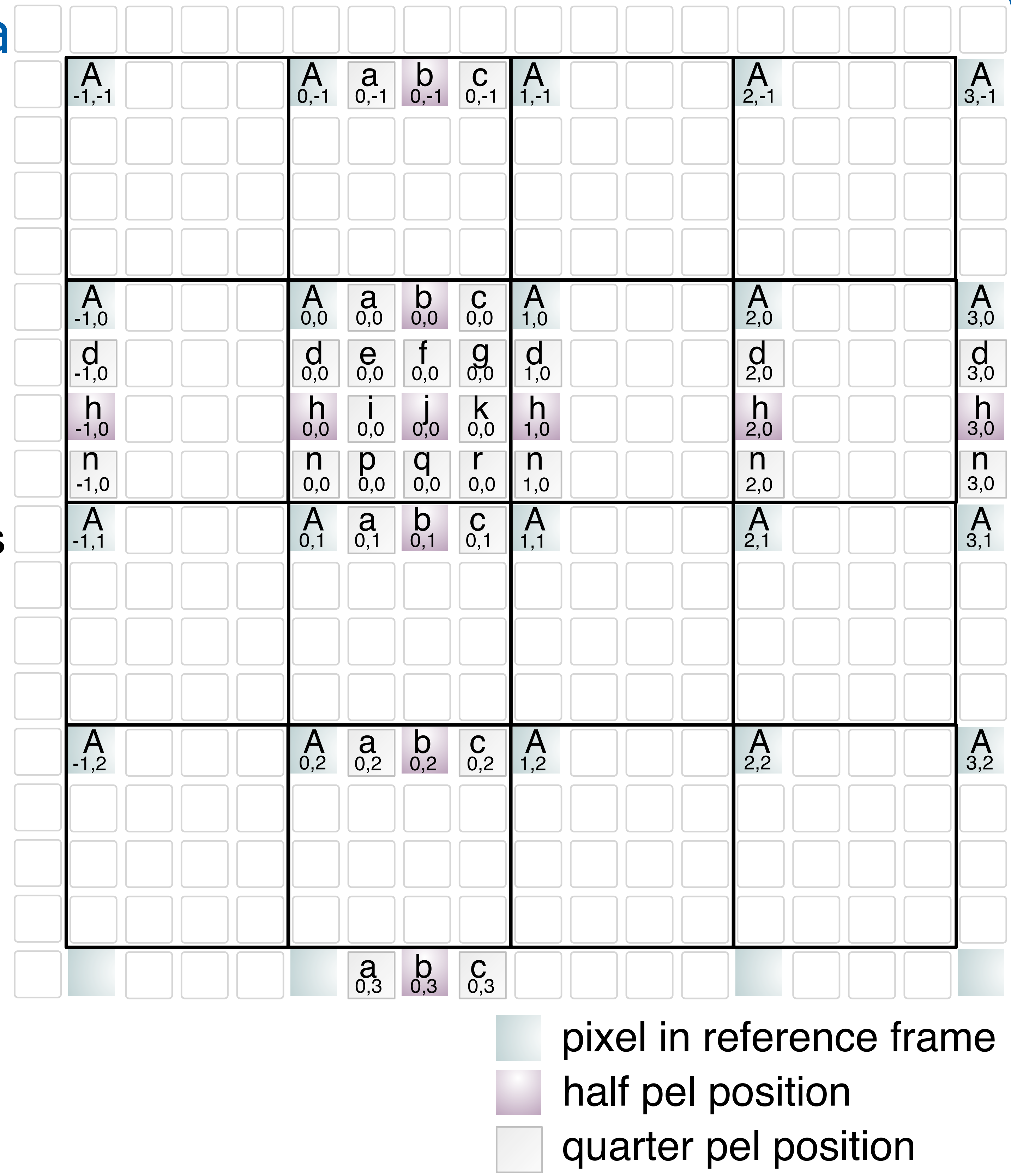
Intra Prediction

- Smoothed predictors from directly adjacent neighbors (left, above the current block)
- Square sizes from 4x4 … 32x32 luma pixels
- Intra_Angular Prediction with 33 different directions of prediction
  - non-uniform angles in directional modes: denser mode coverage for near vertical and near horizontal prediction angles
  - for a block size of NxN, 4N+1 spatially neighboring samples are used
  - bi-linear interpolation with 1/32 sample accuracy
- Intra_Planar
  - four corner reference samples, average of two linear projections
- Intra_DC Prediction
  - average of predictor samples copied to the whole block
- additional boundary smoothing for DC, horizontal, vertical
- by selecting an index from the 3 most probable modes or 5 bit FLC



Motion Compensation - Luma

- Single-stage interpolation process
  - MPEG-4 and H.264 required multiple computation passes
  - HEVC result is available after just two interpolation filter steps (horizontal+vertical)
- Separable horizontal and vertical filters
- 7 filter taps (half-pel positions) or 8 filter taps (quarter pel positions)
- 16 Bit word accuracy sufficient for computation



Motion Compensation - Luma (2)

- Filter Coefficients

index	-3	-2	-1	0	1	2	3	4
hfilter[i]	-1	4	-11	40	40	-11	4	1
qfilter[i]	-1	4	-10	58	17	-5	1	

- First stage (B is bits per pixel)

$$a_{0,j} = \left(\sum_{i=-3...3} A_{i,j} \ \text{qfilter}[i] \right) >> (B-8)$$
$$b_{0,j} = \left(\sum_{i=-3...4} A_{i,j} \ \text{hfilter}[i] \right) >> (B-8)$$
$$c_{0,j} = \left(\sum_{i=-2...4} A_{i,j} \ \text{qfilter}[1-i] \right) >> (B-8)$$
$$d_{0,0} = \left(\sum_{j=-3...3} A_{0,j} \ \text{qfilter}[j] \right) >> (B-8)$$
$$h_{0,0} = \left(\sum_{j=-3...4} A_{0,j} \ \text{hfilter}[j] \right) >> (B-8)$$
$$n_{0,0} = \left(\sum_{j=-2...4} A_{0,j} \ \text{qfilter}[1-j] \right) >> (B-8)$$

- Second Stage

$$e_{0,0} = \left(\sum_{j=-3...3} a_{0,j} \ \text{qfilter}[j] \right) >> 6$$
$$f_{0,0} = \left(\sum_{j=-3...3} b_{0,j} \ \text{qfilter}[j] \right) >> 6$$
$$g_{0,0} = \left(\sum_{j=-3...3} c_{0,j} \ \text{qfilter}[j] \right) >> 6$$
$$i_{0,0} = \left(\sum_{j=-3...4} a_{0,j} \ \text{hfilter}[j] \right) >> 6$$
$$j_{0,0} = \left(\sum_{j=-3...4} b_{0,j} \ \text{hfilter}[j] \right) >> 6$$
$$k_{0,0} = \left(\sum_{j=-3...4} c_{0,j} \ \text{hfilter}[j] \right) >> 6$$
$$p_{0,0} = \left(\sum_{j=-2...4} a_{0,j} \ \text{qfilter}[1-j] \right) >> 6$$
$$q_{0,0} = \left(\sum_{j=-2...4} b_{0,j} \ \text{qfilter}[1-j] \right) >> 6$$
$$r_{0,0} = \left(\sum_{j=-2...4} c_{0,j} \ \text{qfilter}[1-j] \right) >> 6$$

- Third stage: optionally, the interpolation process is followed by weighted prediction (like H.264, only explicit mode supported)
- for bi-directional prediction, interpolated results are added together
- last: rounding/clipping to $[0,2^B\text{-}1]$ is performed

x>>n operator: arithmetic right shift ($x/2^n$)

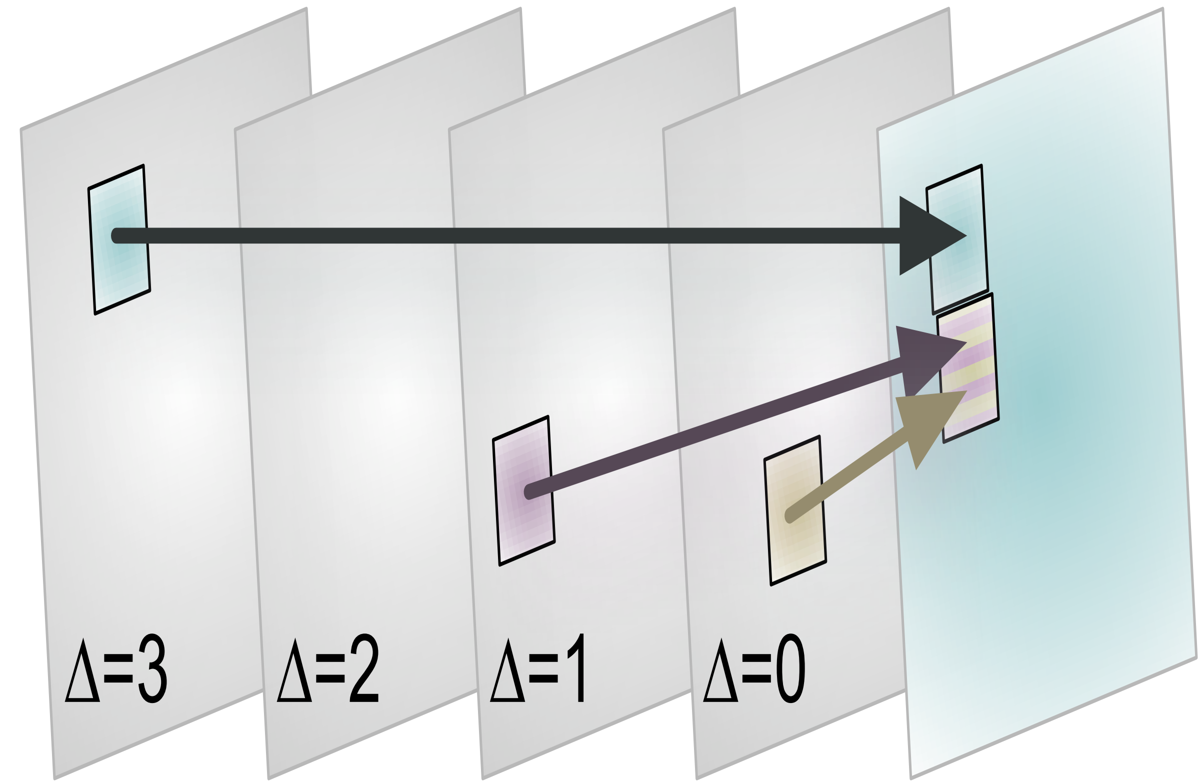
Motion Compensation - Chroma

- Chroma MC is 1/8 pel resolution in YCbCr 4:2:0
- 4 tap FIR filter, indices 5-7 are the mirrored indices 3-1
- horizontal / vertical filtering separately, depending on fractional position
- workflow in analogy to Luma MC

index	-1	0	1	2
filter1[i]	-2	58	10	-2
filter2[i]	-4	54	16	-2
filter3[i]	-6	46	28	-4
filter4[i]	-4	36	36	-4
filter5[i]	-4	28	46	-6
filter6[i]	-2	16	54	-4
filter7[i]	-2	10	58	-2

Motion Compensation - reference frames

- each motion vector is assigned a reference frame index Δ
- reference frames kept on encoder and decoder side
- in bi-directional prediction: two sets of motion vector parameters (list0,list1)
- sorting of reference frames by Picture Order Count (POC)
- ordering transmitted in the slice header (RPS=reference picture set)
- simplified and more robust syntax compared to H.264



Motion Compensation - Merge Mode Motion Inference

- Multiple motion vector prediction candidates in spatial and temporal directions
  - Generalization and extension of the concepts toward DIRECT and SKIP modes in earlier standards
  - Multi-hypothesis motion information
- Temporal motion candidates are stored with a granularity of 16x16 for memory efficiency
- Set of motion vector prediction candidates
  - spatial neighbor candidates
    - based on availability and PU location
    - limit towards the number can be signaled so that only the first candidates are retained
  - temporal candidate from collocated reference picture
    - index explicitly transmitted to decide which reference frame list
  - generated candidates
    - pre-defined list of usual motion vectors, included into merge candidate list for B-Slices
    - zero (0,0) vectors appended to the list for P-Slices
- Candidate list is pruned to remove redundant entries
- For each PU, the index into the candidate list is transmitted
- Similar multi-hypothesis approach to non-merge motion prediction (limited set)

Residual Transform

- Integer DCT approximation for 4x4,8x8,16x16,32x32 transform sizes
  - One-dimensional transform in horizontal and vertical directions, followed by 7 Bit shift, along with 16 Bit clipping to fit all intermediately stored data into 16 Bit
  - Each PU can be either transformed directly as 1 TB or subdivided into 4x4 TBs
- Core transform H, as given in the standard:

$H =$

64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64
90	90	88	85	82	78	73	67	61	54	46	38	31	22	13	4	-4	-13	-22	-31	-38	-46	-54	-61	-67	-73	-78	-82	-85	-88	-90	-90
90	87	80	70	57	43	25	9	-9	-25	-43	-57	-70	-80	-87	-90	-90	-87	-80	-70	-57	-43	-25	-9	9	25	43	57	70	80	87	90
90	82	67	46	22	-4	-31	-54	-73	-85	-90	-88	-78	-61	-38	-13	13	38	61	78	88	90	85	73	54	31	4	-22	-46	-67	-82	-90
89	75	50	18	-18	-50	-75	-89	-89	-75	-50	-18	18	50	75	89	89	75	50	18	-18	-50	-75	-89	-89	-75	-50	-18	18	50	75	89
88	67	31	-13	-54	-82	-90	-78	-46	-4	38	73	90	85	61	22	-22	-61	-85	-90	-73	-38	4	46	78	90	82	54	13	-31	-67	-88
87	57	9	-43	-80	-90	-70	-25	25	70	90	80	43	-9	-57	-87	-87	-57	-9	43	80	90	70	25	-25	-70	-90	-80	-43	9	57	87
85	46	-13	-67	-90	-73	-22	38	82	88	54	-4	-61	-90	-78	-31	31	78	90	61	4	-54	-88	-82	-38	22	73	90	67	13	-46	-85
83	36	-36	-83	-83	-36	36	83	83	36	-36	-83	-83	-36	36	83	83	36	-36	-83	-83	-36	36	83	83	36	-36	-83	-83	-36	36	83
82	22	-54	-90	-61	13	78	85	31	-46	-90	-67	4	73	88	38	-38	-88	-73	-4	67	90	46	-31	-85	-78	-13	61	90	54	-22	-82
80	9	-70	-87	-25	57	90	43	-43	-90	-57	25	87	70	-9	-80	-80	-9	70	87	25	-57	-90	-43	43	90	57	-25	-87	-70	9	80
78	-4	-82	-73	13	85	67	-22	-88	-61	31	90	54	-38	-90	-46	46	90	38	-54	-90	-31	61	88	22	-67	-85	-13	73	82	4	-78
75	-18	-89	-50	50	89	18	-75	-75	18	89	50	-50	-89	-18	75	75	-18	-89	-50	50	89	18	-75	-75	18	89	50	-50	-89	-18	75
73	-31	-90	-22	78	67	-38	-90	-13	82	61	-46	-88	-4	85	54	-54	-85	4	88	46	-61	-82	13	90	38	-67	-78	22	90	31	-73
70	-43	-87	9	90	25	-80	-57	57	80	-25	-90	-9	87	43	-70	-70	43	87	-9	-90	-25	80	57	-57	-80	25	90	9	-87	-43	70
67	-54	-78	38	85	-22	-90	4	90	13	-88	-31	82	46	-73	-61	61	73	-46	-82	31	88	-13	-90	-4	90	22	-85	-38	78	54	-67
64	-64	-64	64	64	-64	-64	64	64	-64	-64	64	64	-64	-64	64	64	-64	-64	64	64	-64	-64	64	64	-64	-64	64	64	-64	-64	64
61	-73	-46	82	31	-88	-13	90	-4	-90	22	85	-38	-78	54	67	-67	-54	78	38	-85	-22	90	4	-90	13	88	-31	-82	46	73	-61
57	-80	-25	90	-9	-87	43	70	-70	-43	87	9	-90	25	80	-57	-57	80	25	-90	9	87	-43	-70	70	43	-87	-9	90	-25	-80	57
54	-85	-4	88	-46	-61	82	13	-90	38	67	-78	-22	90	-31	-73	73	31	-90	22	78	-67	-38	90	-13	-82	61	46	-88	4	85	-54
50	-89	18	75	-75	-18	89	-50	-50	89	-18	-75	75	18	-89	50	50	-89	18	75	-75	-18	89	-50	-50	89	-18	-75	75	18	-89	50
46	-90	38	54	-90	31	61	-88	22	67	-85	13	73	-82	4	78	-78	-4	82	-73	-13	85	-67	-22	88	-61	-31	90	-54	-38	90	-46
43	-90	57	25	-87	70	9	-80	80	-9	-70	87	-25	-57	90	-43	-43	90	-57	-25	87	-70	-9	80	-80	9	70	-87	25	57	-90	43
38	-88	73	-4	-67	90	-46	-31	85	-78	13	61	-90	54	22	-82	82	-22	-54	90	-61	-13	78	-85	31	46	-90	67	4	-73	88	-38
36	-83	83	-36	-36	83	-83	36	36	-83	83	-36	-36	83	-83	36	36	-83	83	-36	-36	83	-83	36	36	-83	83	-36	-36	83	-83	36
31	-78	90	-61	4	54	-88	82	-38	-22	73	-90	67	-13	-46	85	-85	46	13	-67	90	-73	22	38	-82	88	-54	-4	61	-90	78	-31
25	-70	90	-80	43	9	-57	87	-87	57	-9	-43	80	-90	70	-25	-25	70	-90	80	-43	-9	57	-87	87	-57	9	43	-80	90	-70	25
22	-61	85	-90	73	-38	-4	46	-78	90	-82	54	-13	-31	67	-88	88	-67	31	13	-54	82	-90	78	-46	4	38	-73	90	-85	61	-22
18	-50	75	-89	89	-75	50	-18	-18	50	-75	89	-89	75	-50	18	18	-50	75	-89	89	-75	50	-18	-18	50	-75	89	-89	75	-50	18
13	-38	61	-78	88	-90	85	-73	54	-31	4	22	-46	67	-82	90	-90	82	-67	46	-22	-4	31	-54	73	-85	90	-88	78	-61	38	-13
9	-25	43	-57	70	-80	87	-90	90	-87	80	-70	57	-43	25	-9	-9	25	-43	57	-70	80	-87	90	-90	87	-80	70	-57	43	-25	9
4	-13	22	-31	38	-46	54	-61	67	-73	78	-82	85	-88	90	-90	90	-90	88	-85	82	-78	73	-67	61	-54	46	-38	31	-22	13	-4

Residual Transform (2)

- Matrices $H^{(16)}$, $H^{(8)}$, $H^{(4)}$ can be obtained from H by extracting each $2\mathrm{nd}$, $4\mathrm{th}$, $8\mathrm{th}$ row/ column, respectively
- Recursive and partially factored implementation possible instead of straight-forward (and computationally expensive) matrix multiplication

- Mode-dependent alternative transform
  - Only Intra 4x4 transform can be selectively performed by an alternative algorithm
  - Integer DST (discrete sine transform)
  - observation that the prediction error increases with the distance from the border samples used as predictors
  - approx. 1% bit rate reduction in intra-predictive coding

$$H = \begin{vmatrix} 29 & 55 & 74 & 84 \\ 74 & 74 & 0 & -74 \\ 84 & -29 & -74 & 55 \\ 55 & -84 & 74 & -29 \end{vmatrix}$$

Coefficient Scanning, Quantization, Entropy Coding

- Coefficient Scanning, Block coding
  - three different coefficient scanning methods: diagonal, horizontal, vertical

- Quantization
  - Applied transforms are scaled orthonormal DCT/DST
  - Frequency uniform Quantization/Reconstruction
  - QP 0...51 (like H.264), logarithmic step size

- Entropy Coding
  - CABAC (from H.264) as single supported coding method
  - Modified context modeling, significant reduction in context models compared to H.264
  - Higher use of the bypass-mode in CABAC in order to reduce the computational demands
  - Last nonzero coefficient signaling, significance map, sign bit and level encoding concepts similar to H.264
  - Significance map grouped for multiple 4x4 blocks

In-loop Filters

- Two post-processing filter steps in HEVC

- Deblocking Filter (DBF)
  - similar in concept to H.264 deblocking filter
  - low pass filtering of decoded frames for improved visual appearance and <u>improved coding efficiency</u>, reduction of block artifacts
  - significant reduction in complexity compared to H.264 approach
  - multi-processing friendly

- Sample adaptive offset (SAO)
  - applied after the Deblocking Filter
  - Suppression of „banding artifacts“ from strong quantization as well as „ringing artifacts“ from quantization errors of high frequency components
  - Signaling of SAO parameters flexible, ranging from one parameter specification (over the whole picture) down to a fine granularity on CU level

Deblocking Filter

- Applied on TU and PU boundaries (not necessarily the same)
- 8x8 sample grid for luma and chroma (H.264: 4x4/2x2)
- 3 different strengths for luma
  - 0 no filtering
  - 1 regular filtering for motion compensated blocks (practically the same rules as in H.264 apply)
  - 2 strong Intra filtering if any neighboring block is Intra
- Chroma shares the decisions, yet only two modes (on,off) are defined

- HEVC processing order of DBF is horizontal filtering (vertical edges) first over the whole picture, followed by vertical filtering (horizontal edges)
  - inherently multi-processing friendly

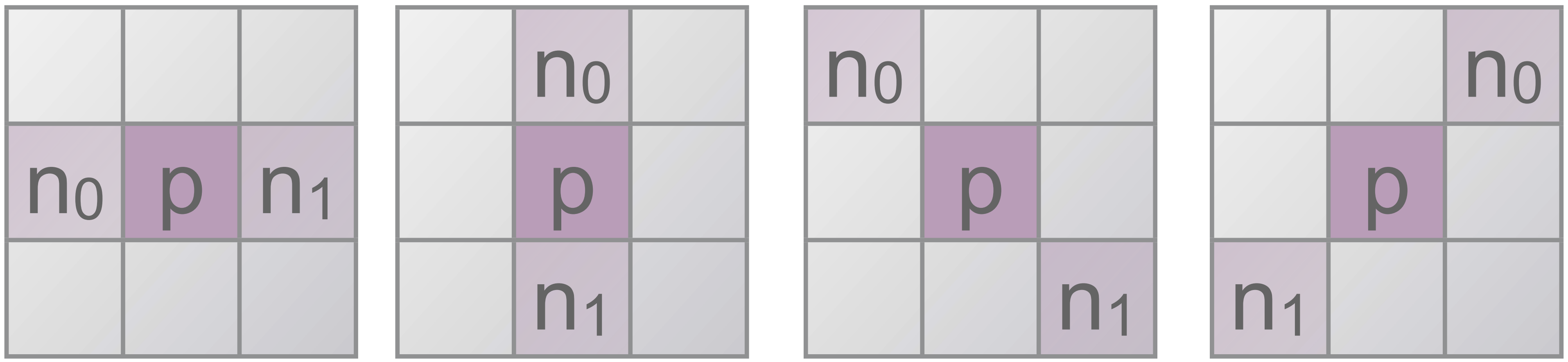
DBF Impact: 176x144, QP36, GOP-Length 32, 30 kBit/s

- DBF off

- DBF on

Sample adaptive offset

- Four gradient patterns in SAO: current pixel p, neighboring pixels n_0,n_1

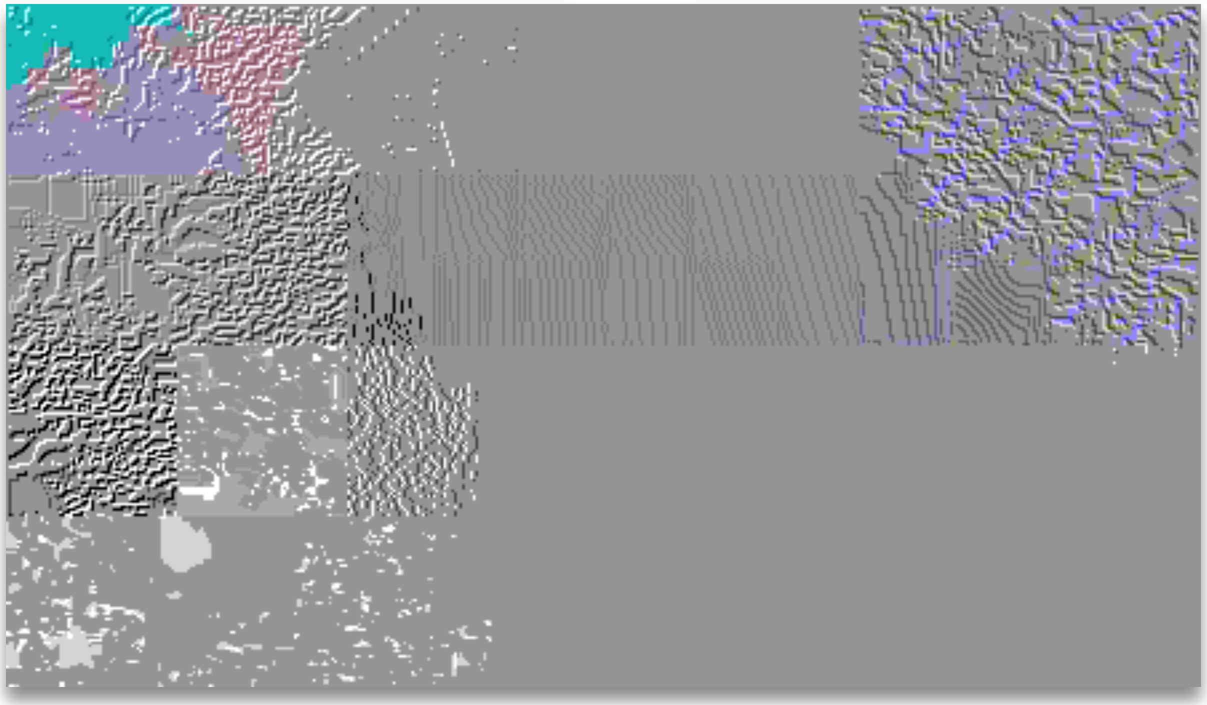


- Syntax element sao_type_idx controls the SAO operation for the current coding tree blocks (CTB, one luma or chroma component of a CTU), either
  - 0 = off
  - 1 = band offset type
    - central band, side band
    - offset depends on sample amplitude (amplitude range divided into 32 bands, sample values of four consecutive bands are modified as band offsets)
  - 2 = edge offset type
    - additional parameter: direction in which the gradient strengths are to be evaluated
    - gradient evaluation and classification done in encoder and decoder
    - encoder sends the appropriate offsets to the decoder for each CTB (positive values only for 1,2, negative values only for 3,4)
- Offsets are added to considered pixels within a coding tree block (luma/chroma) via look-up tables

SAO Impact: 448x256, QP40, GOP-Length 32, 80 kBit/s

- SAO off

- SAO on



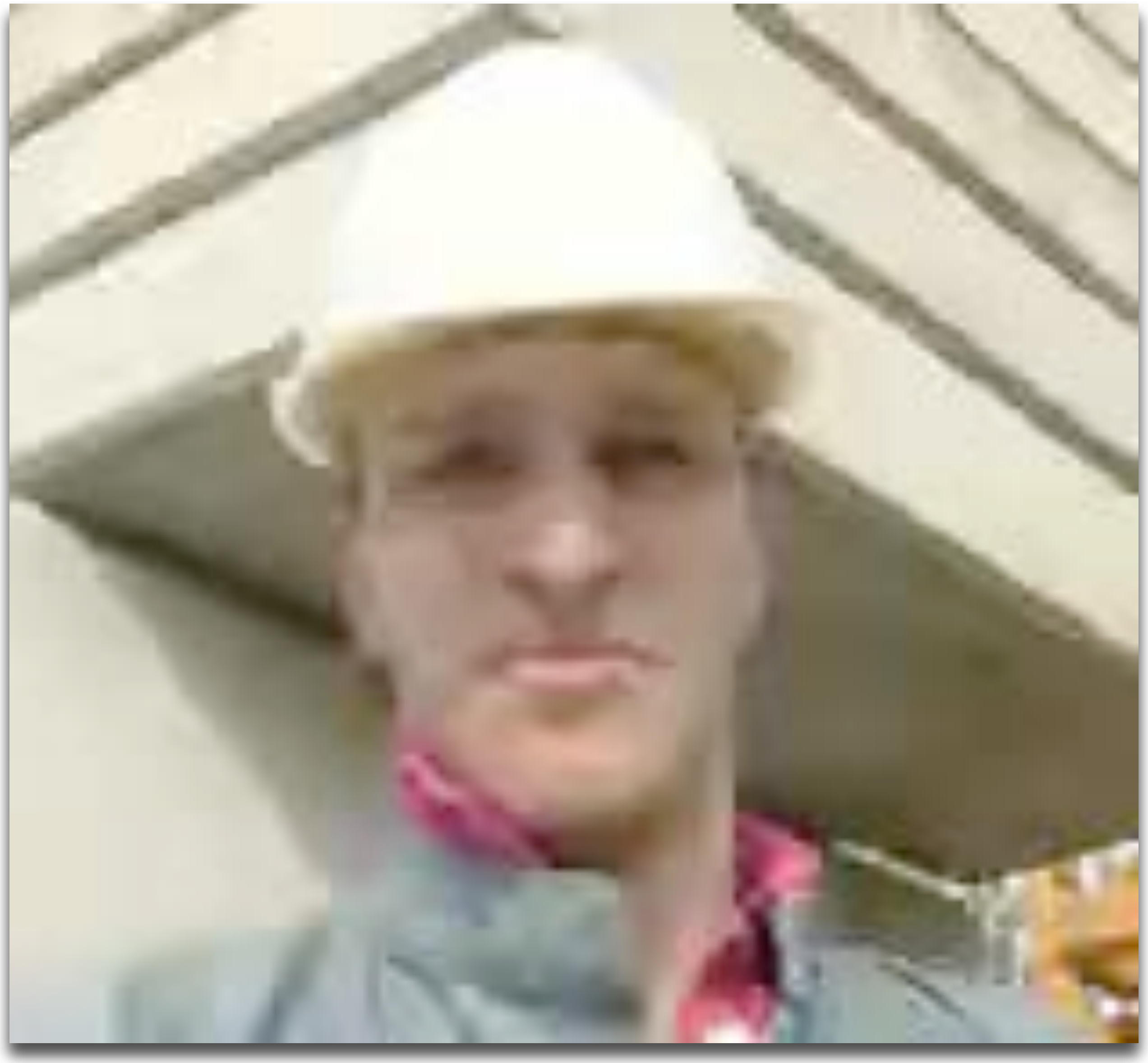
- Contrast enhanced difference image

Further Development

- RXext - Range Extensions
  - Higher pixel depth >10 Bits / sample
  - Additional chroma formats (4:2:2,4:4:4,4:0:0)
  - Additional Channels (e.g. Alpha)
- Scalability Extensions (SHVC)
  - temporal scalability is already part of HEVC v1
  - spatial and SNR scalability
  - multi-loop coding framework (requires decoding of current and all lower layers)
- 3D Video Extensions
  - MV-HEVC (direct sibling to H.264 MVC)
  - Multiview HEVC with modified Block-Level Tools
    - higher compression efficiency than MV-HEVC, backwards compatible to HEVC v1, yet new tools for additional views
- Hybrid Architectures
  - Legacy Base Layer (H.264, H.262) with HEVC enhancement layer

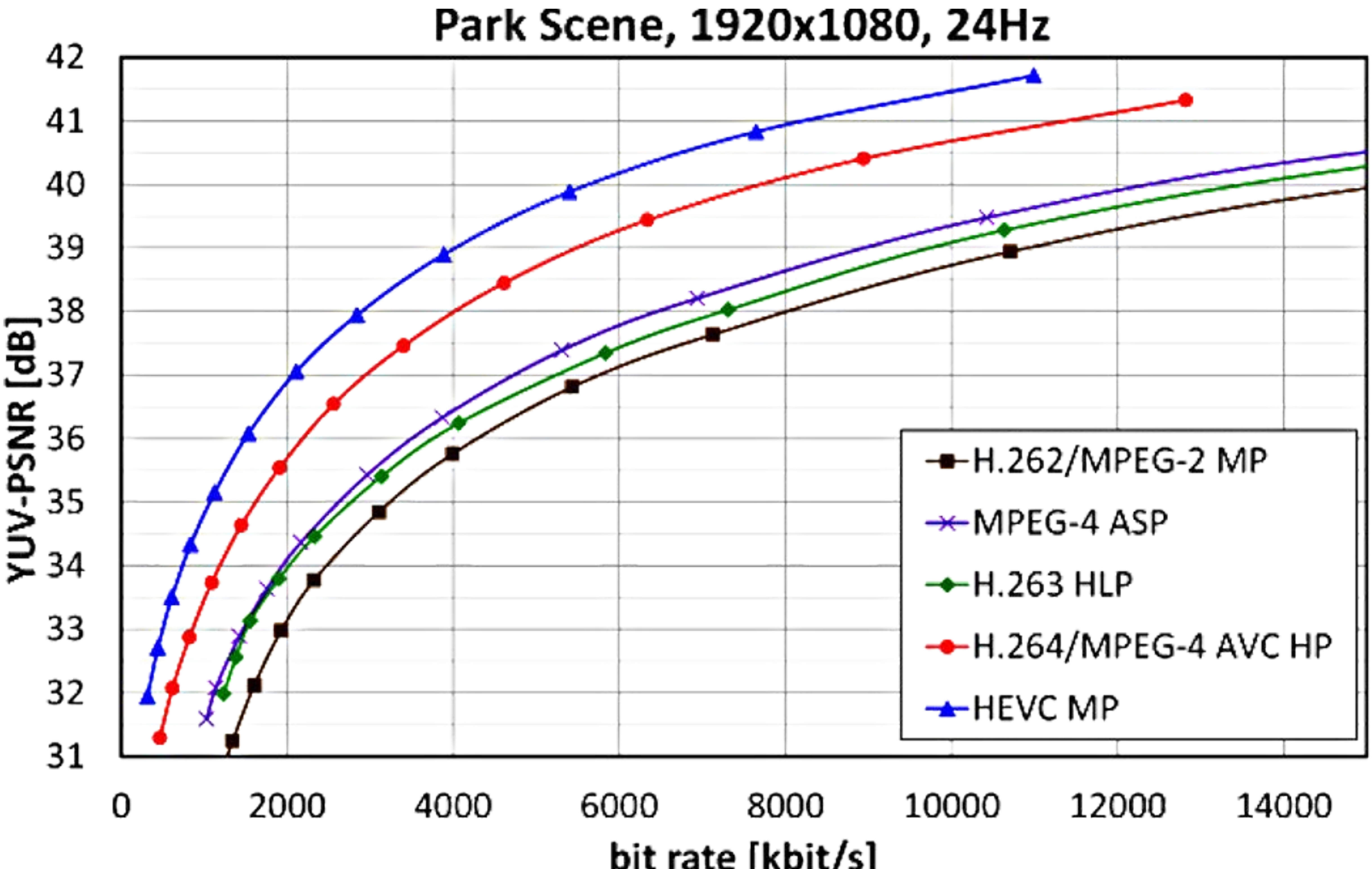
Image Quality H.264 vs. H.265

- Foreman QCIF, first picture (Intra)



- H.264, 9768 Bits, Y-PSNR 30.8 dB (JM12.4)
- H.265, 9208 Bits, Y-PSNR 32.4 dB, (HM16.6)

Comparison of PSNR H.265 ↔ MPEG-2/4,H.264 1080p



Literature

- [SOHW12] G. J. Sullivan, W.J. Han, T. Wiegand, Overview of the High Efficiency Video Coding (HEVC) Standard, IEEE Transactions on Curcuits and Systems for Video Technology, Vol. 22, 1649–1668, Dec 2012
- [SBCOSV13] G. J. Sullivan, J. M. Boyce, Y. Chen, J.R. Ohm, C. A. Segall, A. Vetro, Standardized Extensions of High Efficiency Video Coding (HEVC), IEEE Journal of selected topics in signal processing, Vol. 7, No. 6, 1001-1007, Dec 2013

- Reference software: https://hevc.hhi.fraunhofer.de/
- Free implementation: http://x265.org