HEVC

High Efficiency Video Coding

Jakub Stankowski
Chair of Multimedia Telecommunications and Microelectronics
Poznan University of Technology

Video Compression Standards

- MPEG video compression standards:
  - MPEG-1 Part 2 (Video) – 1993
  - **MPEG-2 Part 2 (Video) / H.262 – 1996**
  - MPEG-4 Part 2 (Visual) (MPEG-4 ASP) – 1999
  - **MPEG-4 Part 10 (Advanced Video Coding) / H.264 – 2003**
  - **MPEG-H Part 2 (High Efficiency Video Coding) / H.265 – 2013**

- Large number of MPEG-2 and MPEG-4 ASP class technologies
- AVC class technologies:
  - Microsoft / SMPTE VC-1
  - AVS
  - VP8/VP9
- Competition for HEVC?

Requirements

- Higher compresion efficiency (compared to AVC)
- Support for UltraHD TV:
  - 4k ,8k resolution
  - wide color gammut
  - higher frame rate (up to 120 fps)
- Support for parallel encoding/decoding
- Avoid some AVC flaws
  - CABAC decoding bottleneck
  - entropy coding dichotomy
  - deblocking filter complexity
  - lack of efficient parallel coding tools

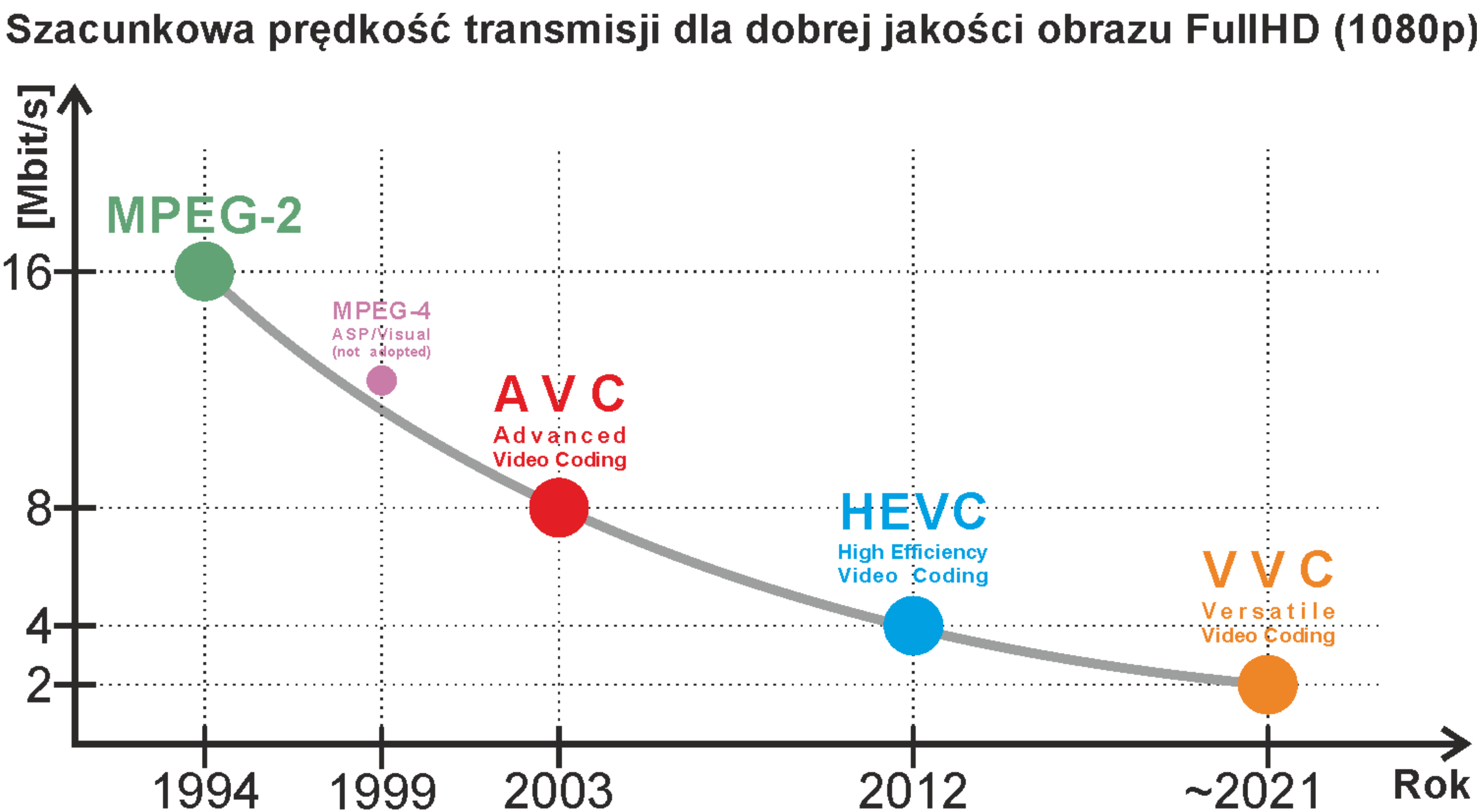
HEVC standardization process

- Requirements definition (2007 – 2010)
- Call for Proposals (Jan 2010)
- Competition phase (Jan 2010 – April 2010)
- Cooperation phase (April 2010 – Jan 2013)
- Committe Draft (CD) (Feb 2012)
- Draft International Standard (DIS) (July 2012)
- Final Draft of International Standard (FDIS) (Jan 2013)
- International Standard (April 2013)

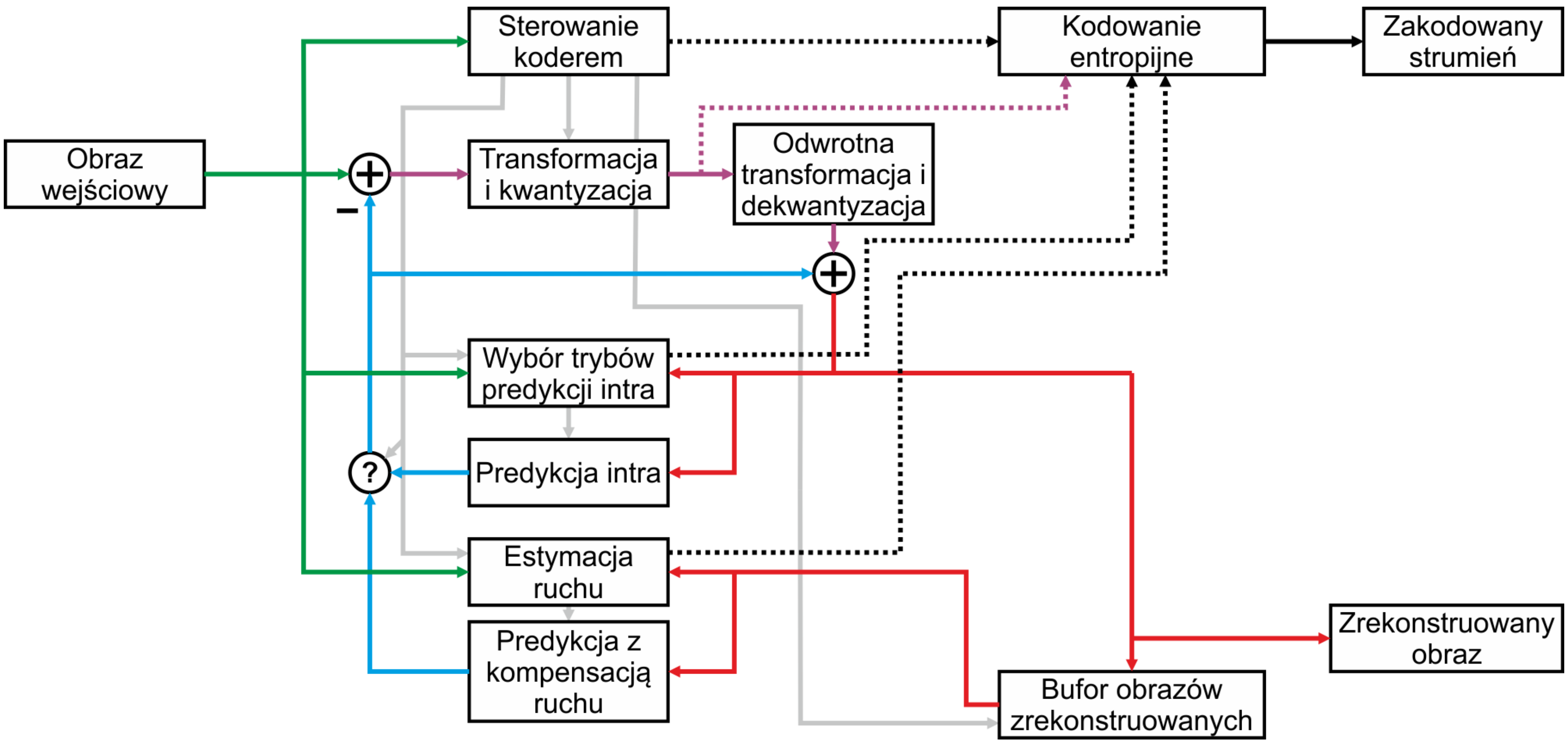
- ISO/IEC standard:
  ISO/IEC 23008-2 Information technology - High Efficiency Coding and Media Delivery in Heterogeneous Environments (MPEG-H) - Part 2 (High Efficiency Video Coding)
- ITU-T standard:
  ITU-T Rec. H.265

Rule of thumb

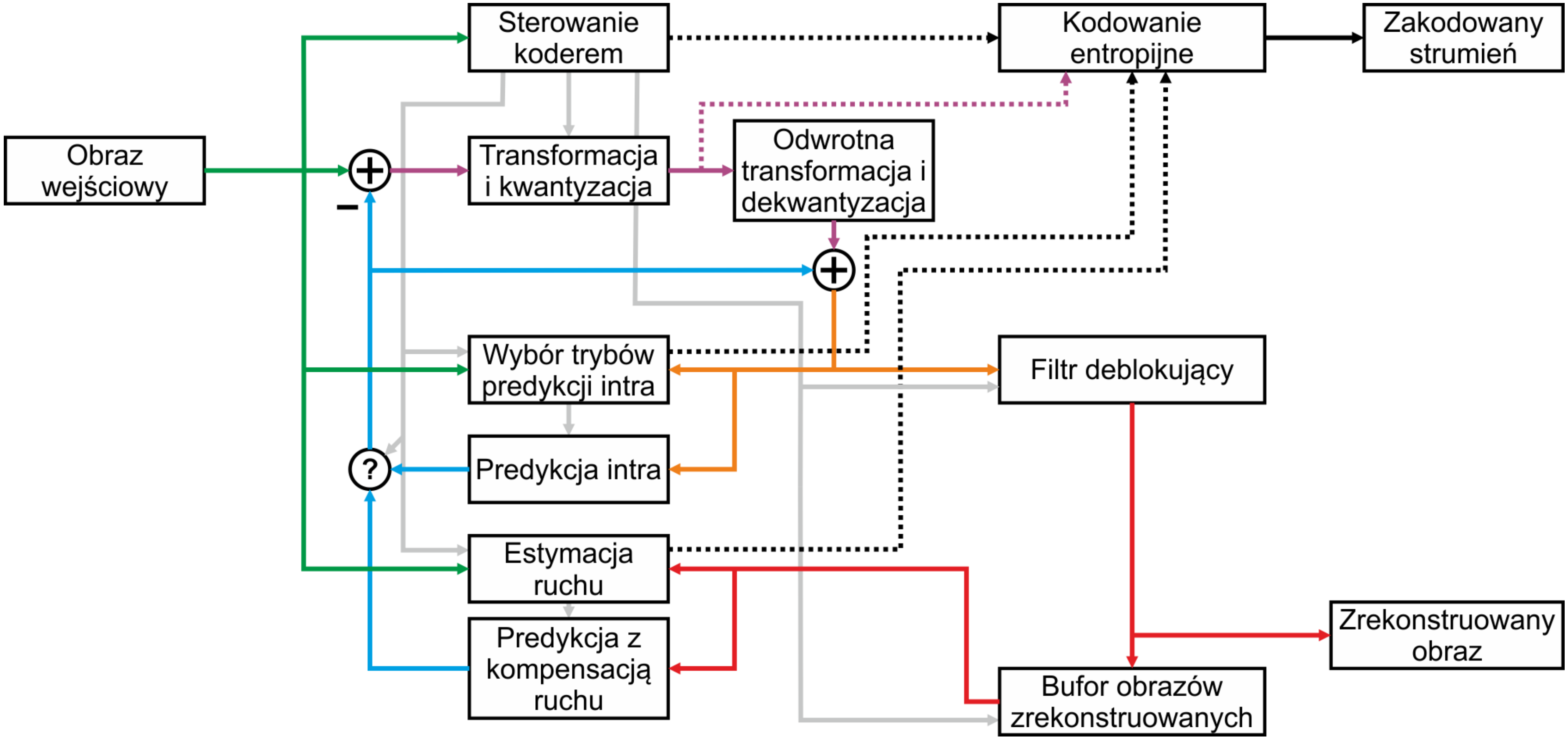
~2x compression efficiency improvement at each generation



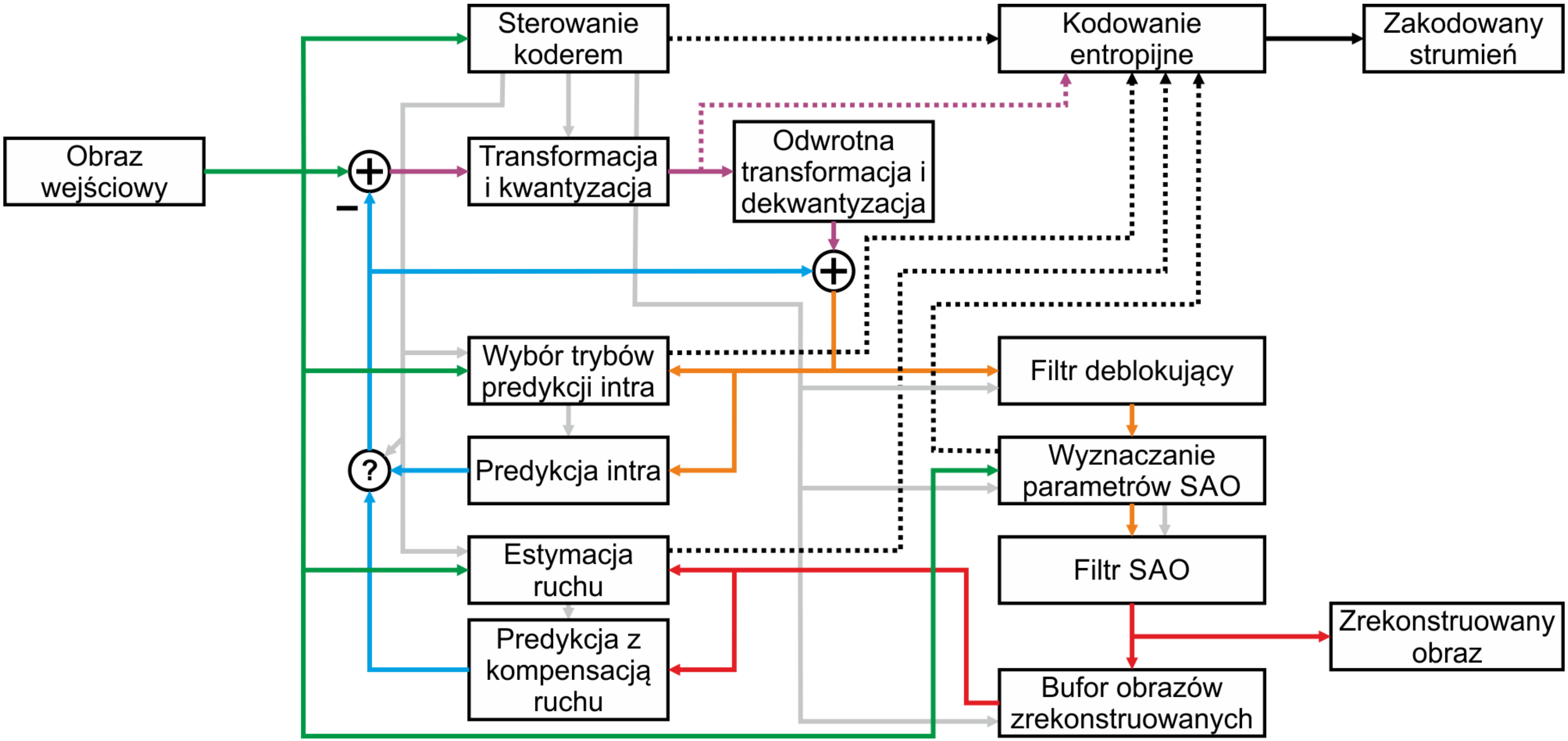
Encoder block diagram MPEG2



Encoder block diagram AVC

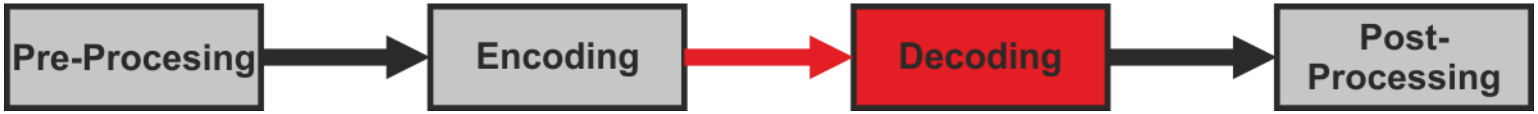


Encoder block diagram MPEG2



Area of standardization

- Bitstream format
- Decoding process
- Bitstream constrains
  - Profile
    (set of coding tools, bit depth)
  - Level, Tier
    (required decoder performance, bandwidth, etc.)



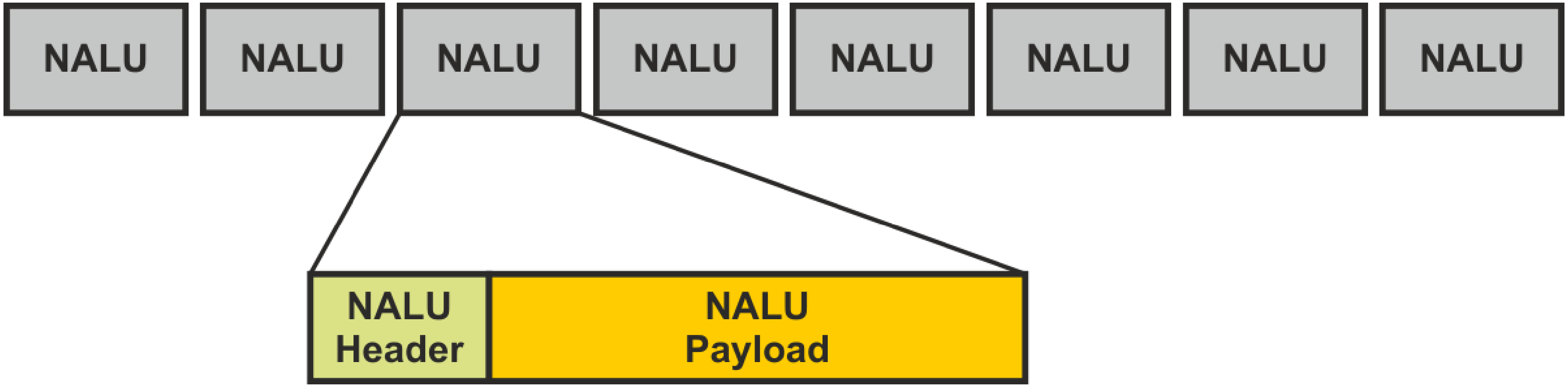
- **Encoder design is NOT standardized**
- Encoder shall only produce conformant bitstream

Levels

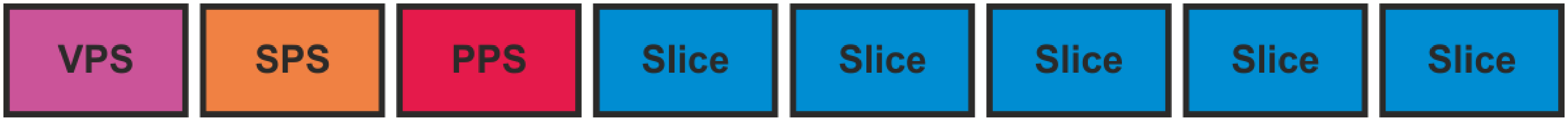
Level	Max luma picture size (samples)	Max CPB size MaxCPB (k bits)		Max slice segments per picture	Max # of tile rows	Max # of tile columns	Example picture resolution @ highest frame rate (Max DPB size)
		Main tier	High tier				
1	36 864	350	-	16	1	1	128×96@33.7 (6), 176×144@15.0 (6)
2	122 880	1 500	-	16	1	1	176×144@100.0 (16), 352×288@30.0 (6)
2.1	245 760	3 000	-	20	1	1	352×288@60.0 (12), 640×360@30.0 (6)
3	552 960	6 000	-	30	2	2	640×360@67.5 (12), 720×576@37.5 (8), 960×540@30.0 (6)
3.1	983 040	10 000	-	40	3	3	720×576@75.0 (12), 960×540@60.0 (8), 1280×720@33.7 (6)
4	2 228 224	12 000	30 000	75	5	5	1,280×720@68.0 (12), 1,920×1,080@32.0 (6) 2,048×1,080@30.0 (6)
4.1	2 228 224	20 000	50 000	75	5	5	1,280×720@136.0 (12), 1,920×1,080@64.0 (6) 2,048×1,080@60.0 (6)
5	8 912 896	25 000	100 000	200	11	10	1,920×1,080@128.0 (16), 3,840×2,160@32.0 (6) 4,096×2,160@30.0 (6)
5.1	8 912 896	40 000	160 000	200	11	10	1,920×1,080@256.0 (16), 3,840×2,160@64.0 (6) 4,096×2,160@60.0 (6)
5.2	8 912 896	60 000	240 000	200	11	10	1,920×1,080@300.0 (16), 3,840×2,160@128.0 (6) 4,096×2,160@120.0 (6)
6	35 651 584	60 000	240 000	600	22	20	3,840×2,160@128.0 (16), 7,680×4,320@32.0 (6) 8,192×4,320@30.0 (6)
6.1	35 651 584	120 000	480 000	600	22	20	3,840×2,160@256.0 (16), 7,680×4,320@64.0 (6) 8,192×4,320@60.0 (6)
6.2	35 651 584	240 000	800 000	600	22	20	3,840×2,160@300.0 (16), 7,680×4,320@128.0 (6) 8,192×4,320@120.0 (6)

Network Abstraction Layer (NAL)

- Coded video data organized into NAL units
- NAL Unit
  - consist of header and payload
  - two types of NAL Units:
    - VCL (Video Coding Layer)
    - Non-VCL
  - header contains NALU type identifier



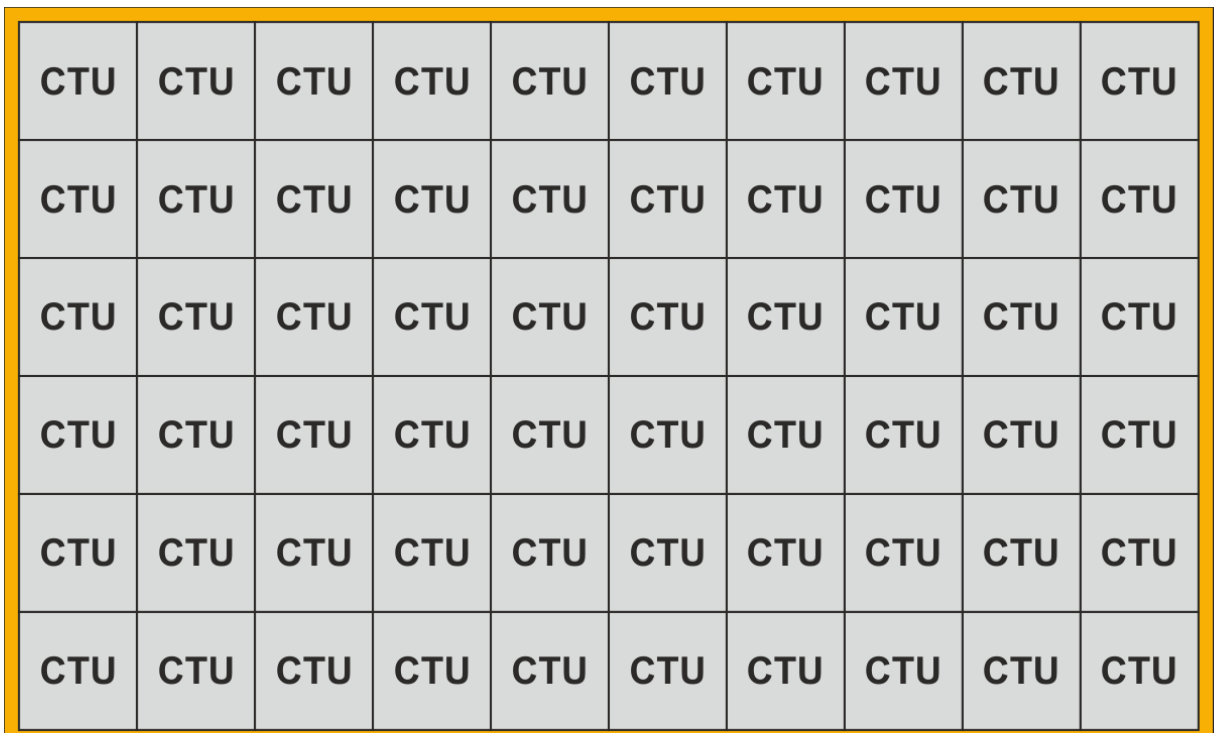
Bitstream elements



- Non-VCL NAL Units:
  - Video Parameter Set (VPS)
  - Sequence Parameter Set (SPS)
  - Picture Parameter Set (PPS)
- VCL NAL Units:
  - Slice of encoded picture

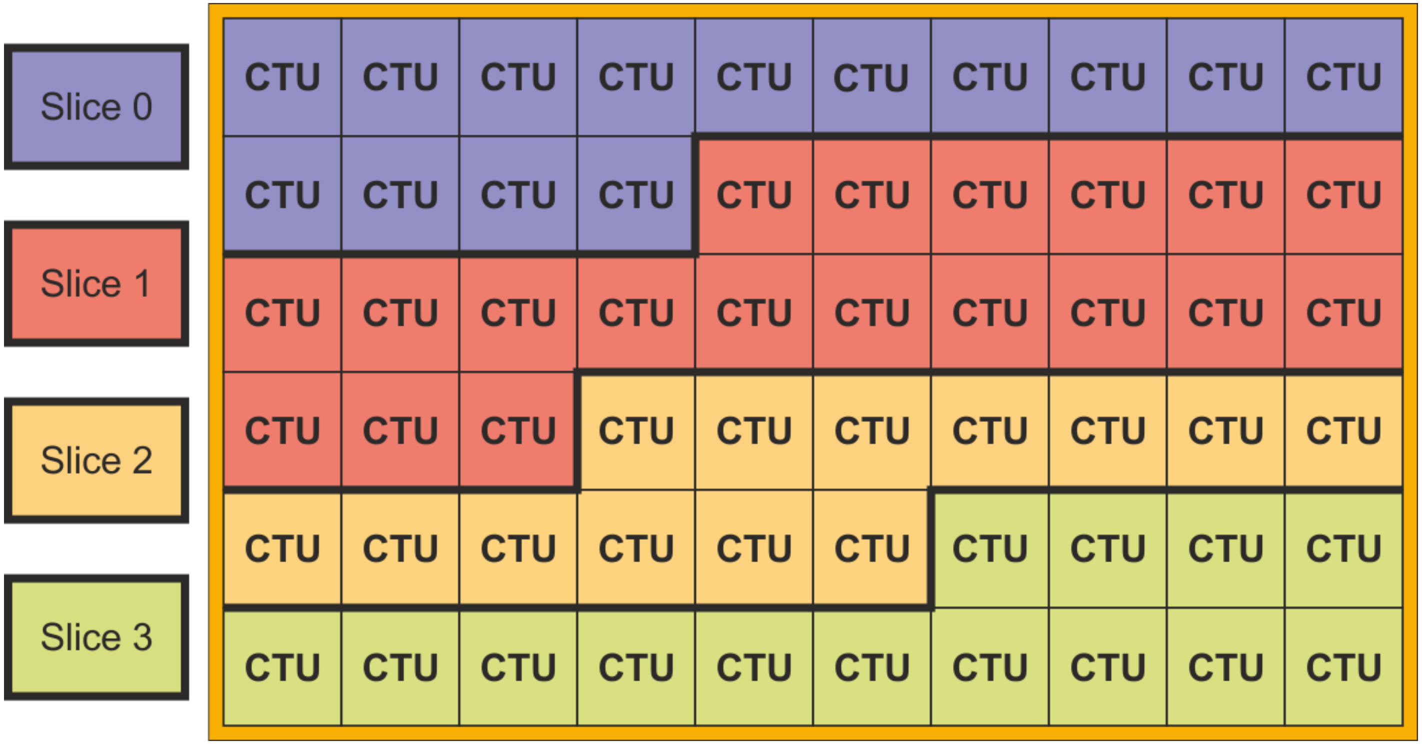
Picture division – CTUs

- Coded picture is partitioned into Coding Tree Units (CTUs)
- CTU corresponds to Macroblock in HEVC
- CTUs are transmitted in raster scan order
- CTU size – **64x64**, 32x32 or 16x16 luma samples

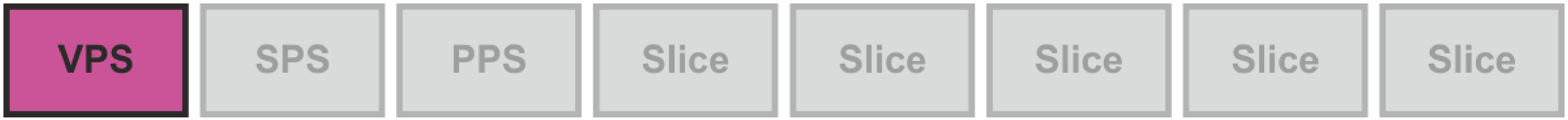


Picture division – Slices

- Picture can be encoded using one or multiple Slices
- Each CTU has to be allocated into exactly one Slice
- Each Slice contain integer number of CTUs
- Slices contain consecutive CTUs in raster scan order

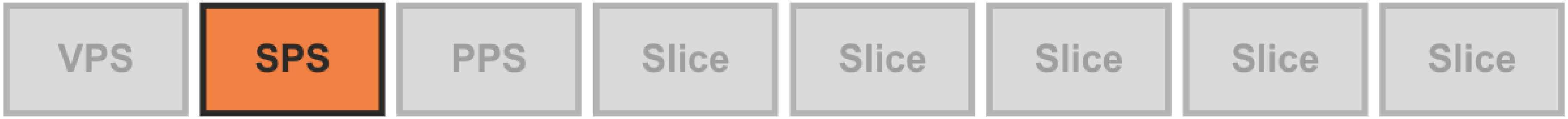


Video Parameter Set



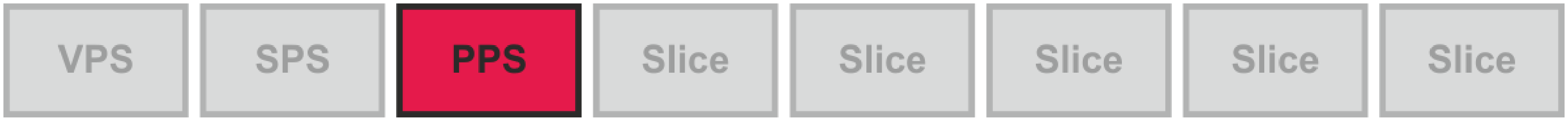
- Contains:
  - High level metadata
  - Overal characteristics of coded video sequences
  - Dependeces between temporal sublayers
  - HRD parameters
  - Profile, Tier, Level identifier

Sequence Parameter Set



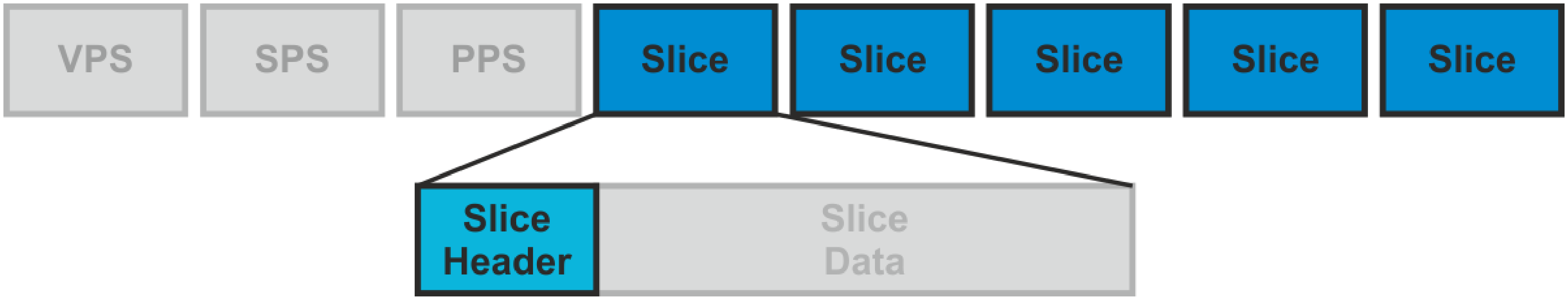
- Contains:
  - Profile, Tier, Level identifier
  - Picture width/height
  - Luma/Chroma bit depth
  - Max/Min Coding Unit size
  - Max/Min Transform Unit size
  - Scaling lists (Scaling Matrices)
  - Reference picture sets
  - VUI Parameters (video metadata)
  - Compresion tools on/off flags

Picture Parameter Set



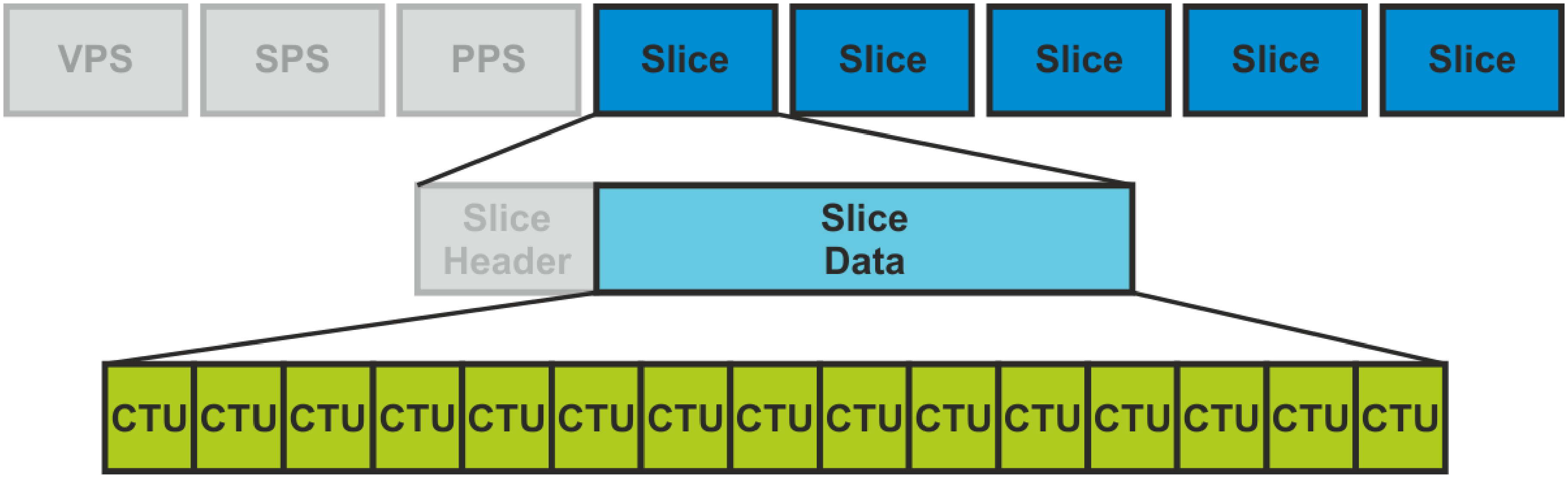
- Contains:
  - Initial QP value
  - Compresion tools on/off flags
  - Tiles/Wavefront parameters
  - Deblocking filter parameters

Slice Header



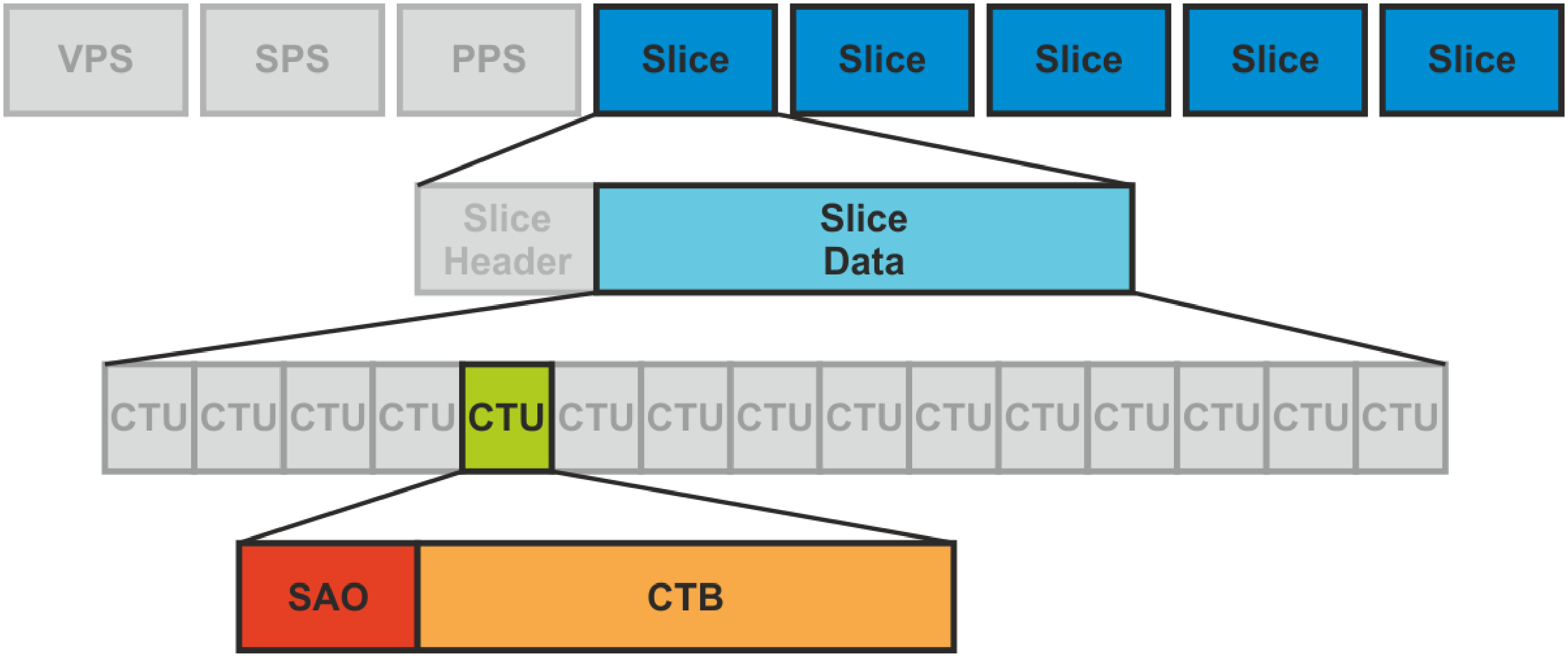
- Contains:
  - Slice type identifier
  - Picture Order Count
  - Reference Picture Set Identifier or New Reference Picture Set
  - Compresion tools on/off flags
  - Compression tools settings
  - Slice QP
  - Deblocking filter parameters
  - Weighted prediction parameters

Slice Data



- Slice data consist of Coding Tree Units (CTU)
- CTU corresponds to Macroblock in HEVC
- CTU size – **64x64**, 32x32 or 16x16 luma samples (CTU size is selected in SPS for all slices in sequence)

Coding Tree Unit



- Sample Adaptive Offset (SAO) – HEVC new inloop tool
- Coding Tree Block (CTB) – contains one or more Coding Units(CU) partitioned using quadtree structures

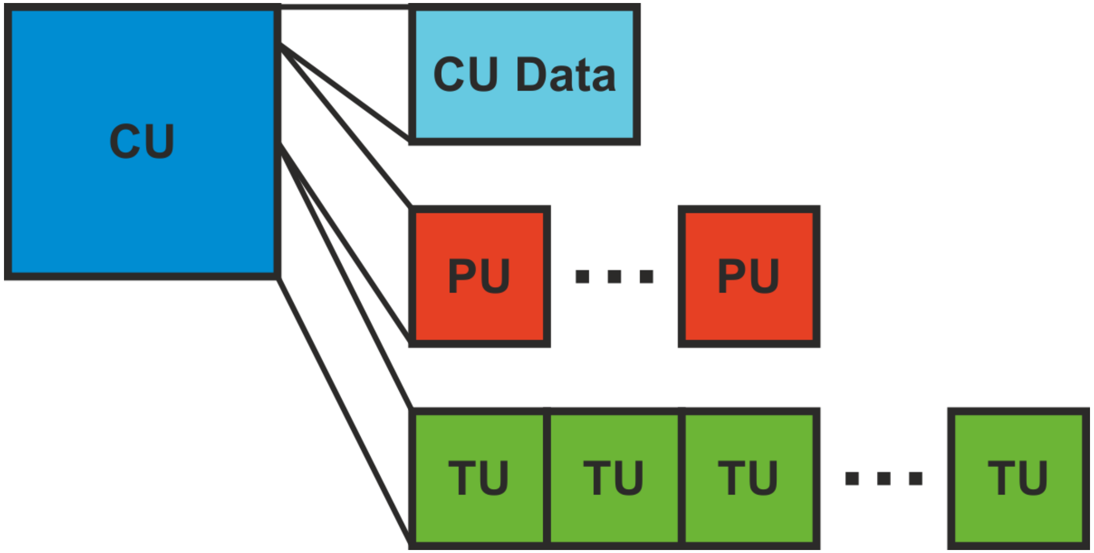
Slice types

- **I-slice** – slice that contains blocks encoded using:
  - intra prediction only
- **P-slice** – slice that contains blocks encoded using:
  - uni-directional inter (motion-compensated) prediction
  - intra prediction
- **B-slice** – slice that contains blocks encoded using:
  - bi-directional inter (motion-compensated) prediction
  - uni-directional inter (motion-compensated) prediction
  - intra prediction
- Coding efficiency / complexity:
  I-slice < P-slice < B-slice

Major HEVC units

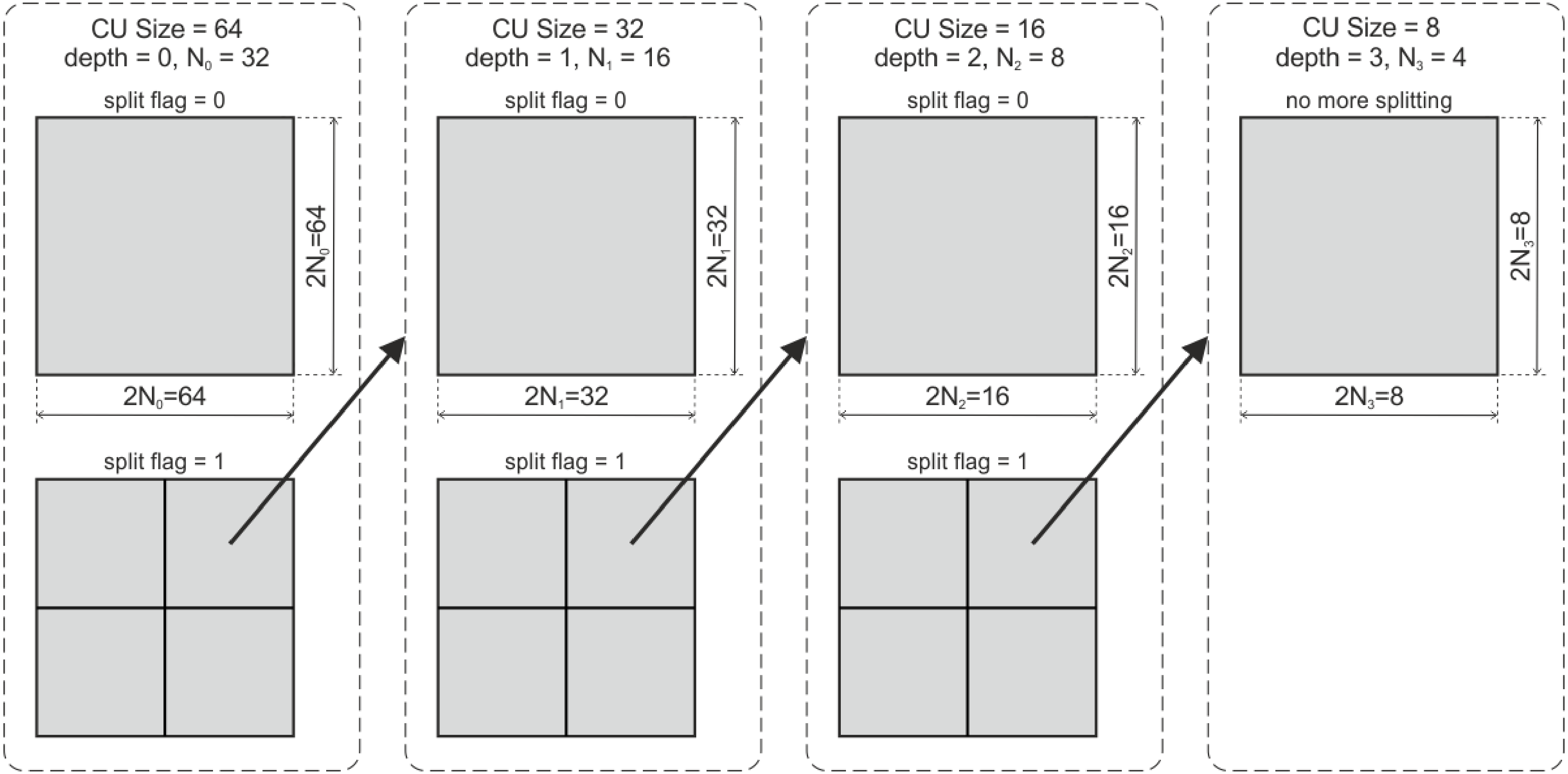
Encoded CTB data is partitioned into:

- Coding Unit (CU)
  - Prediction mode (Intra/Inter/Skip)
  - Partition mode
  - Contain coresponding Prediction Unit(s) and Transtorn Unit(s)
- Prediction Unit (PU)
  - prediction data
- Transform Unit (TU)
  - residual data



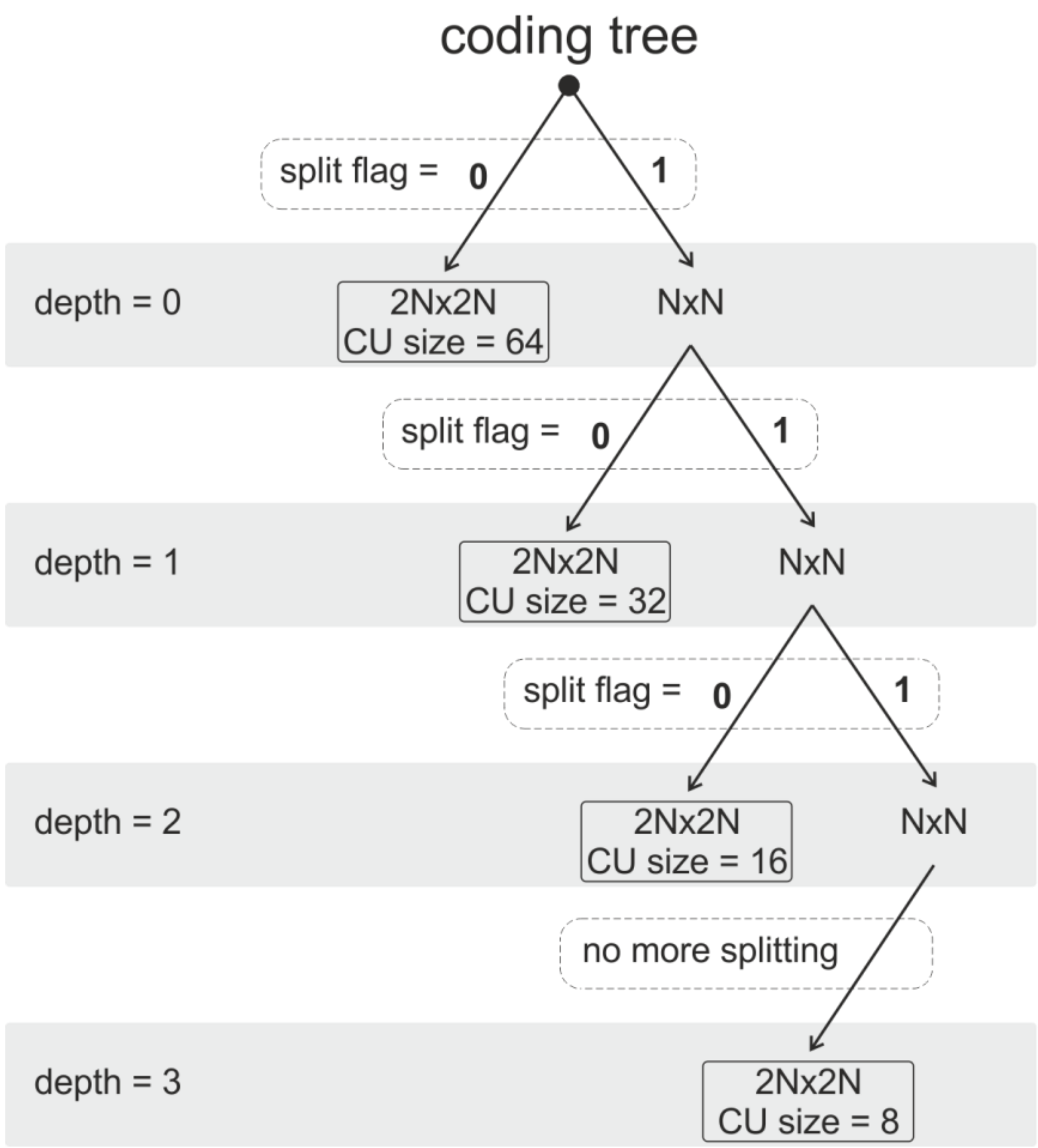
Coding Units

- Encoder can select size of coding units by sending split flags in the Coding Tree



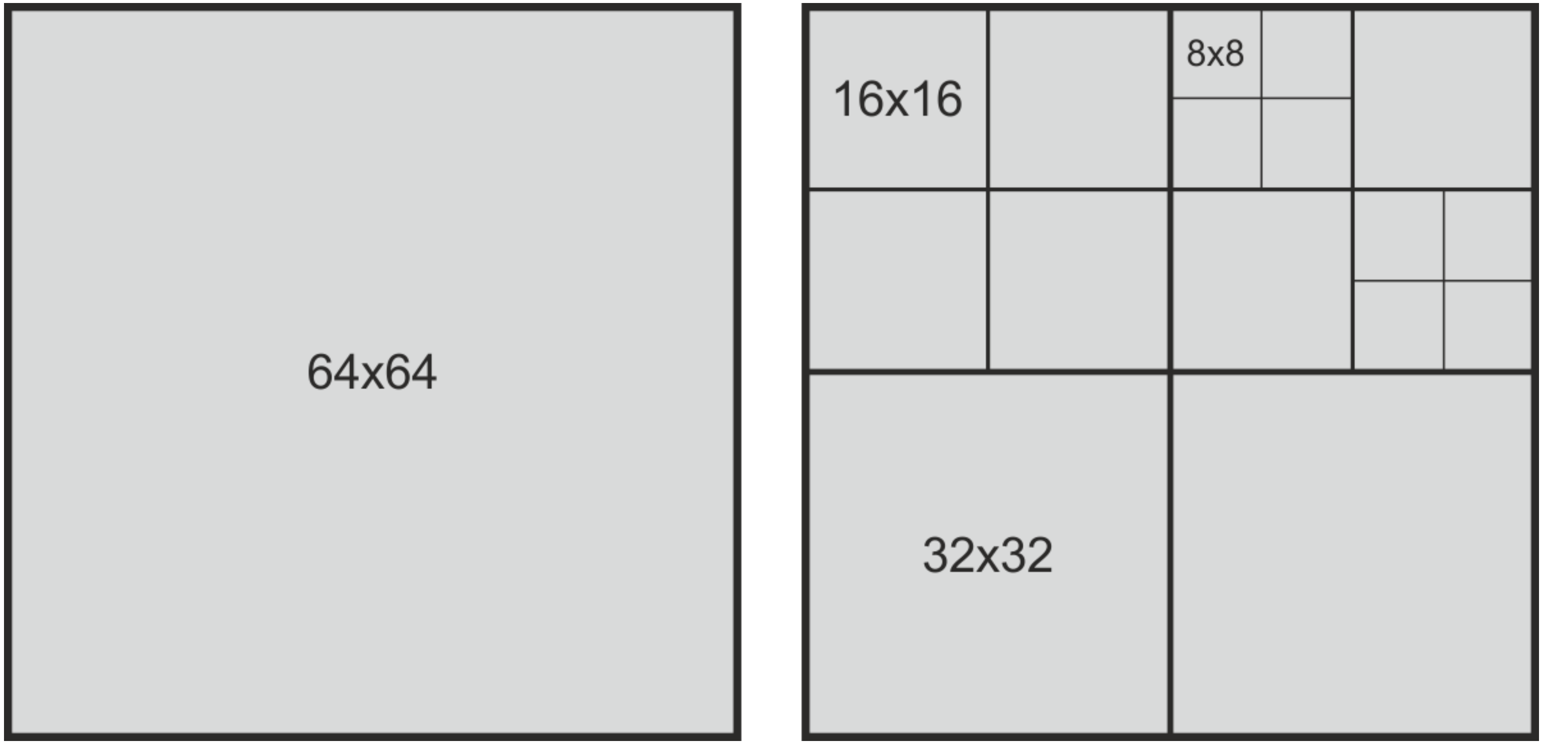
Coding Tree

- Coding Tree Block (CTB) – the root of coding tree
- Coding tree contains one or more Coding Units(CU)
- CUs are partitioned using quadtree structures
- Each CU larger than Smalest CU (SCU) could be splitted into four sub-CUs



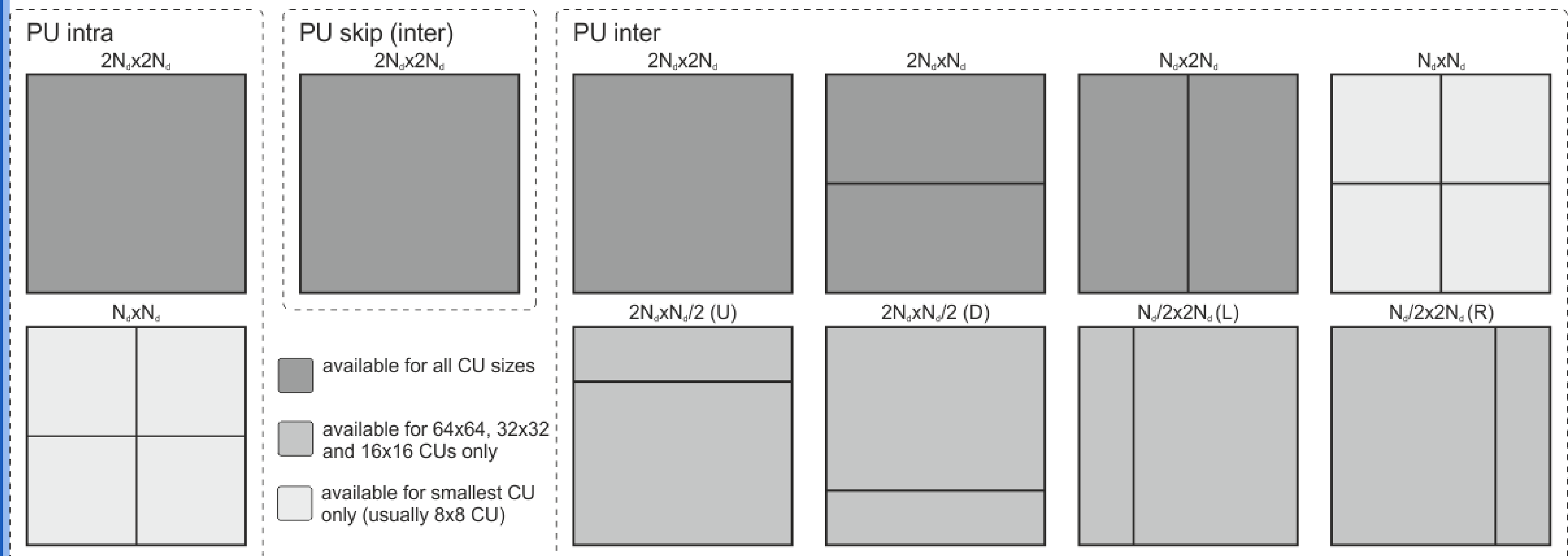
Example: division of CTU into CUs

- Encoder can adjust the CUs size to the image characteristics



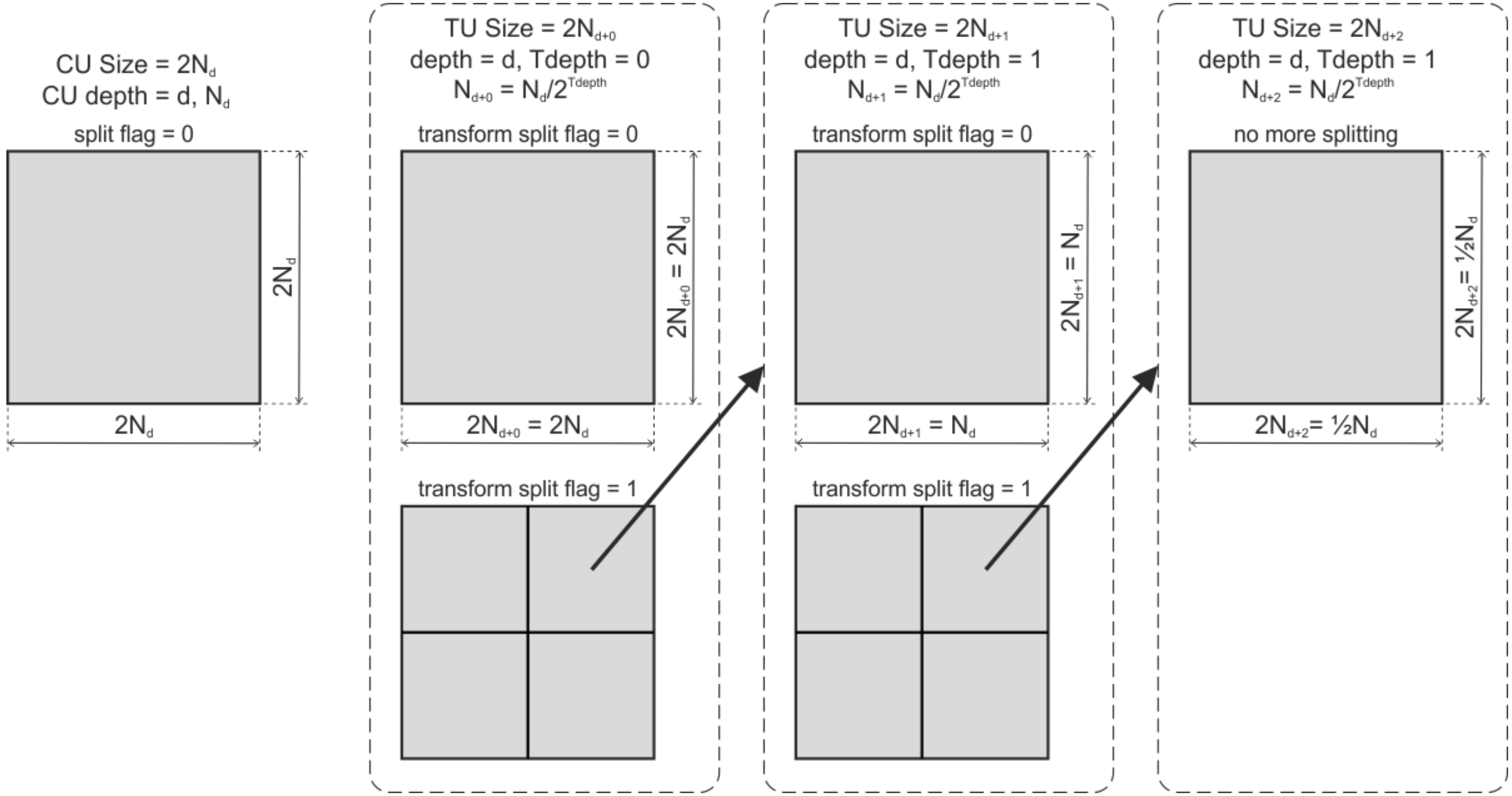
Prediction Unit partitioning

- CU contain one or more PUs
- The set of available partitioning modes depends on prediction type (intra/inter) and CU size
- Exception:
  - 4x4 inter prediction PUs are prohibited
  - 8x4 and 4x8 inter prediction PU are restricted to uni-directional

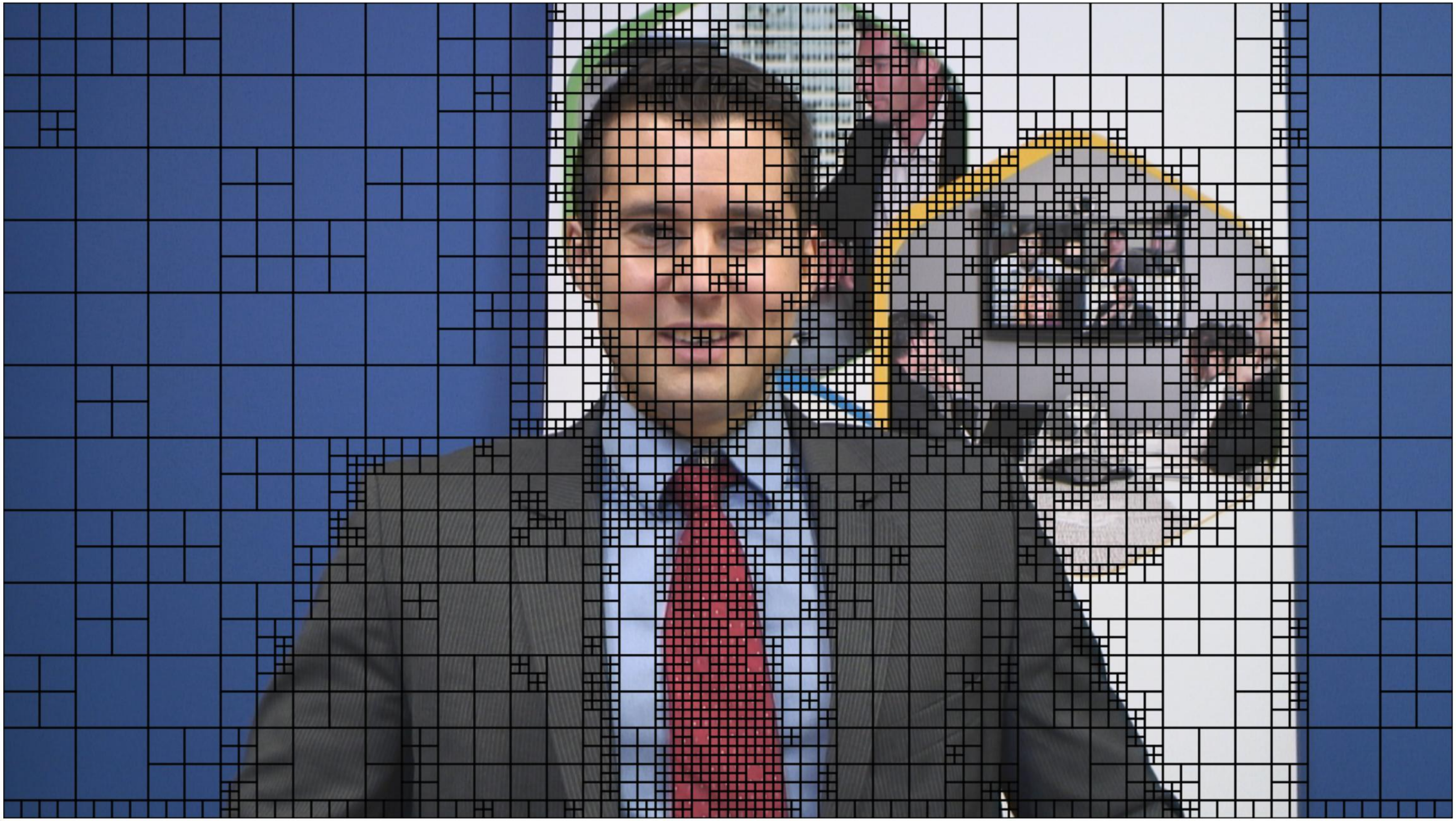


Residual Quadtree (RQT)

- Prediction error (residual signal) is represented as residual quadtree (RQT)
- RQT contains one or more Transform Units (TU)
- Transform Units can be splitted in similar way as CUs

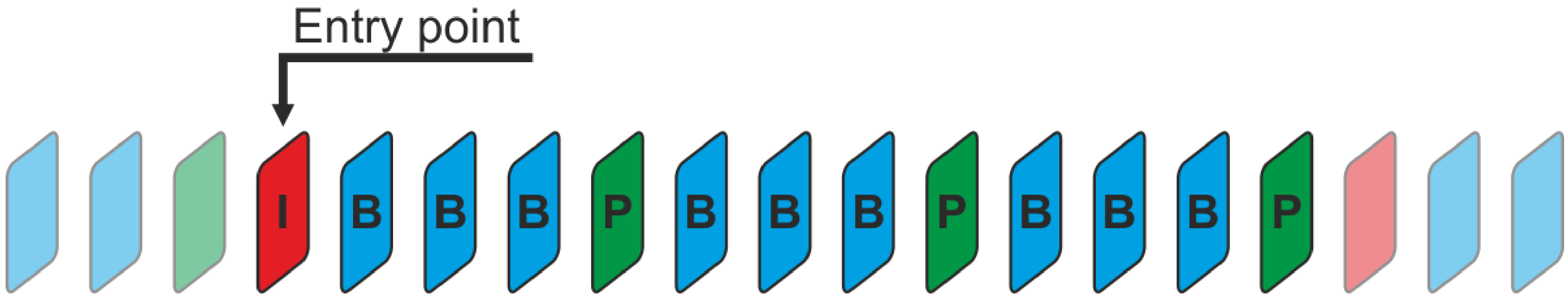


Picture division



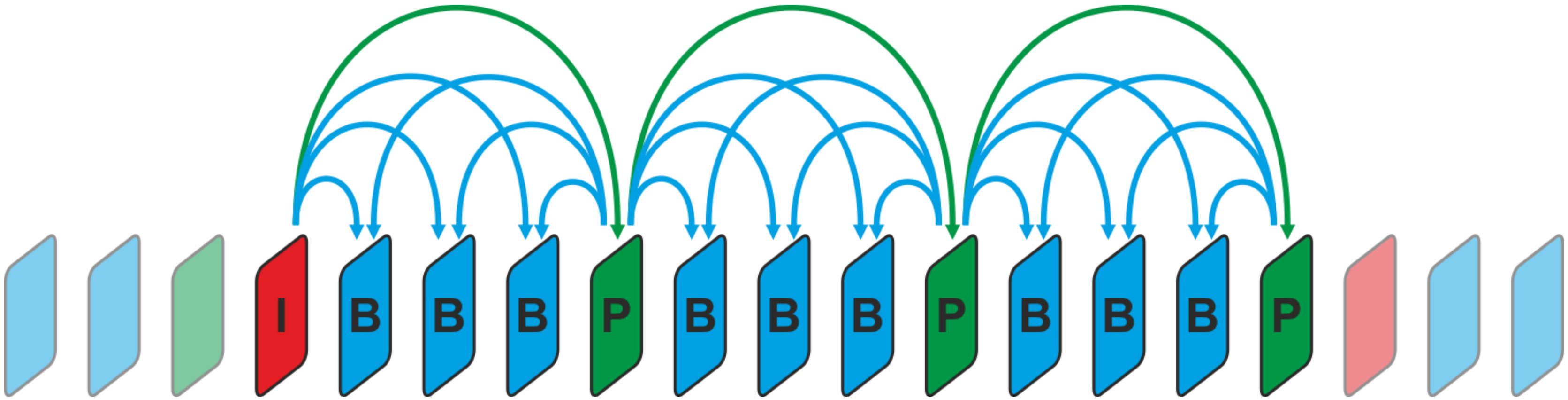
Group of pictures (GOP)

- GOP – a group of successive pictures within a coded video stream
- Contains entry point (I-picture)
- Picture types:
  - **I-picture** – picture encoded using **I-slices**
  - **P-picture** – picture encoded using **P-slices**
  - **B-picture** – picture encoded using **B-slices**



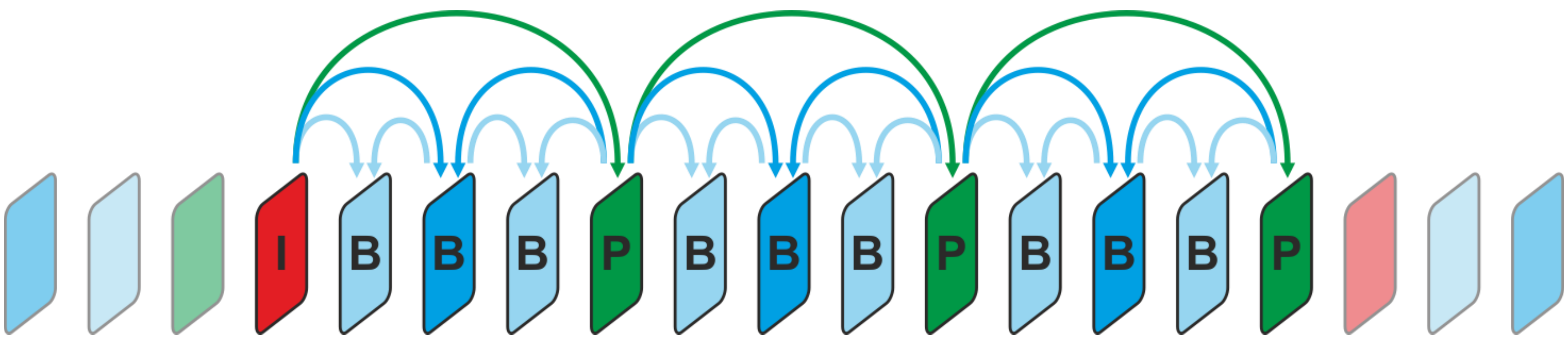
Classic and Hierarchical GOP

Classic GOP:



Hierarchical GOP:

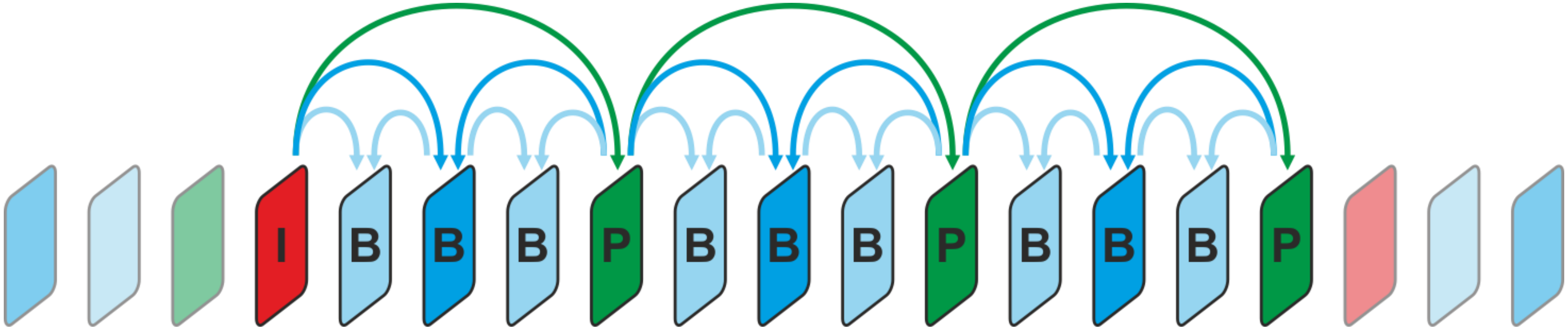
higher compression efficiency + temporal scalability



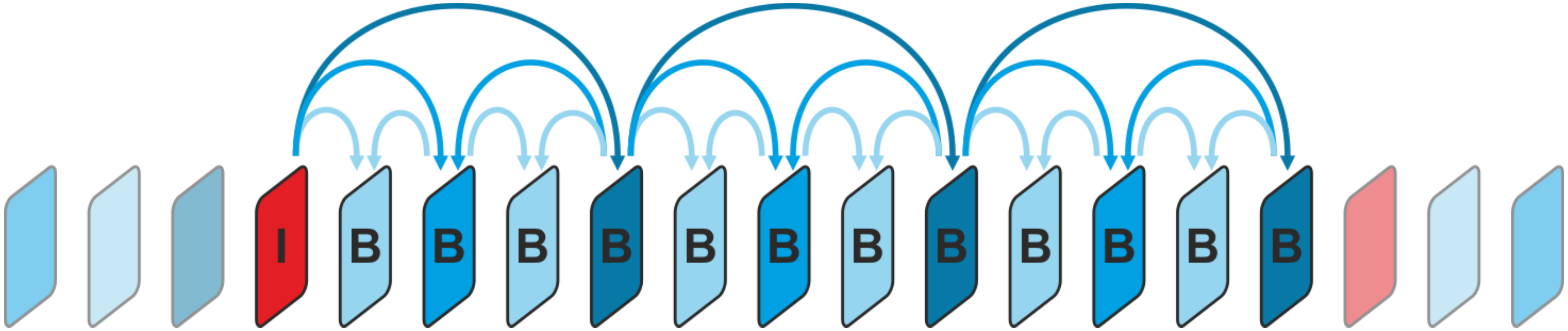
Generalized P and B (GPB)

P-pictures for prediction from backward in time only

B-pictures for prediction from backward and forward in time

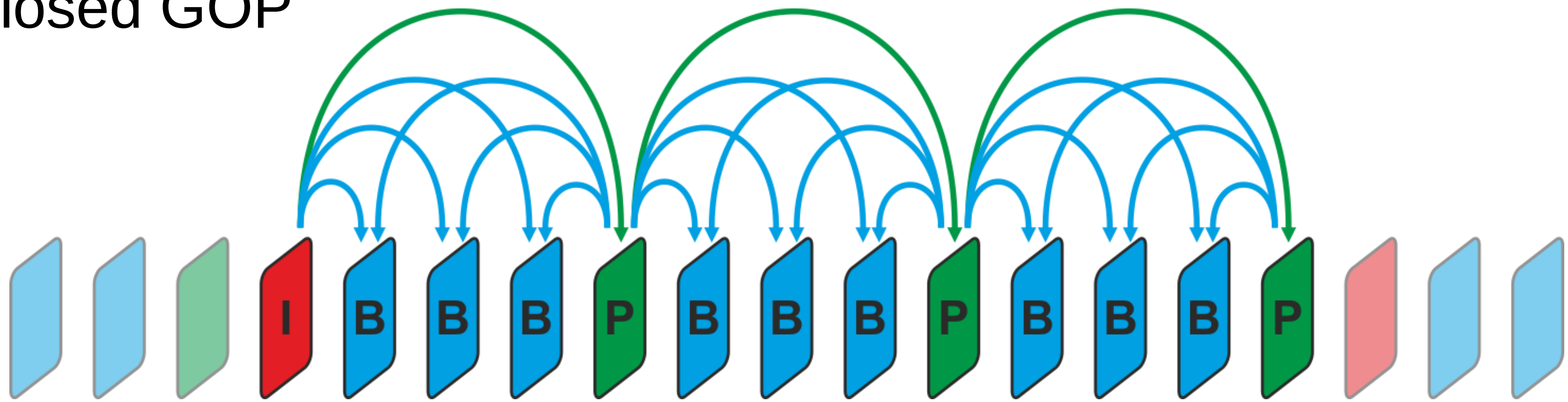


B pictures for prediction from backward only or backward and forward in time



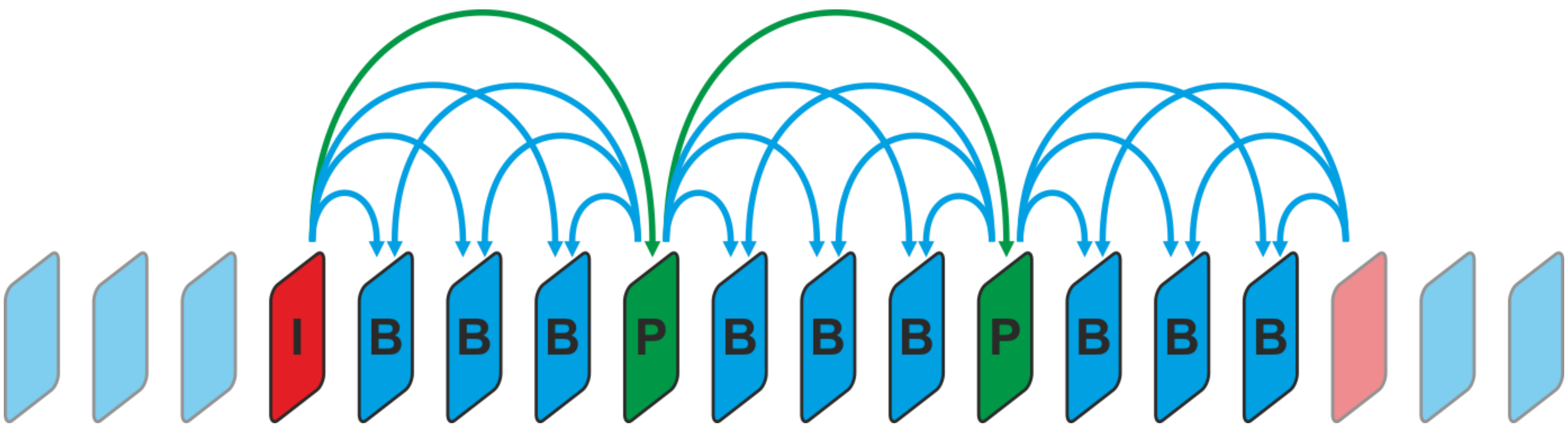
Closed and open GOP

Closed GOP



Open GOP

B-pictures use I-picture from next GOP prediction source



Entry point signalization

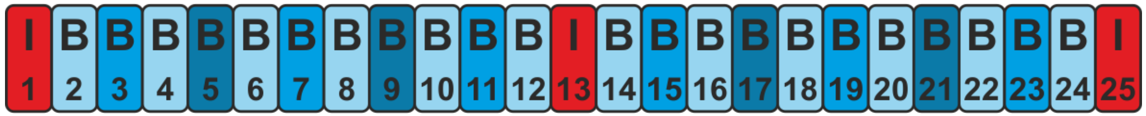
- In AVC entry point is signaled by special type of I-picture: an Instantaneous Decoding Refresh (IDR) picture
- IDR picture definition:

  An IDR picture causes the decoding process to mark all reference pictures as "unused for reference" immediately after the decoding of the IDR picture. All coded pictures that follow an IDR picture in decoding order can be decoded without inter prediction from any picture that precedes the IDR picture in decoding order.

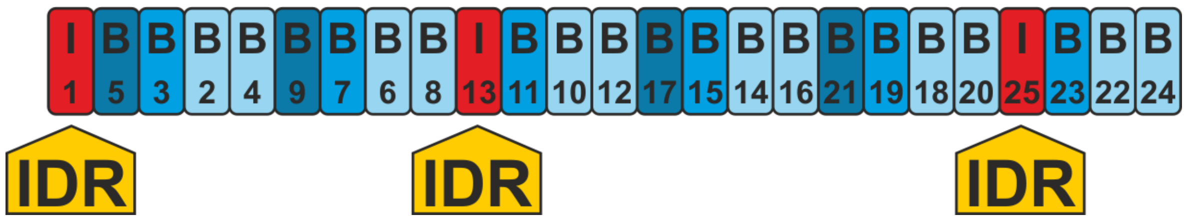
- IDR picture forces decoder to clear its reference picture buffer

Open GOP using IDR pictures

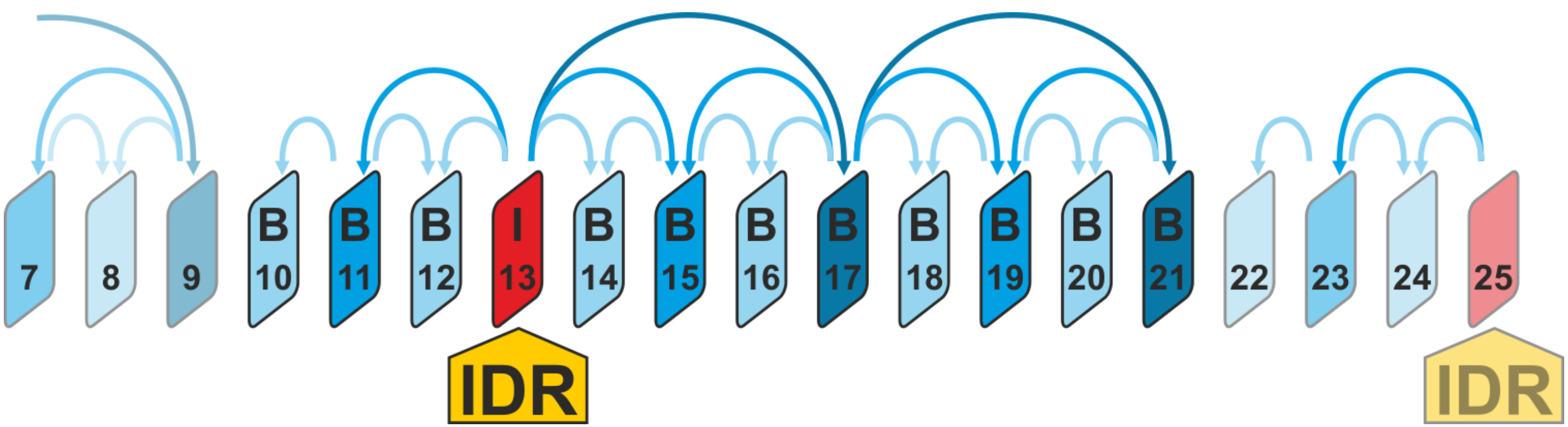
Picture display order:



Picture coding order:



GOP shape and posible inter prediction sources:



Entry point signalization (once again)

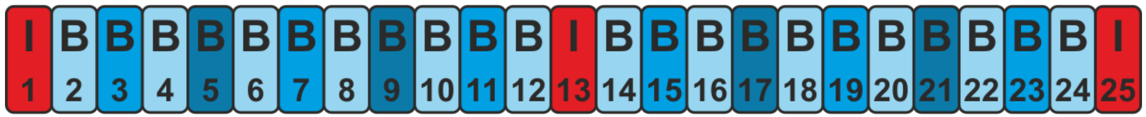
- In HEVC entry point is signaled by three types of I-pictures:
  - Instantaneous Decoding Refresh (IDR) picture
  - Clean Random Acces (CRA) picture
  - Broken Link Acces (BLA) picture

- CRA picture definition:

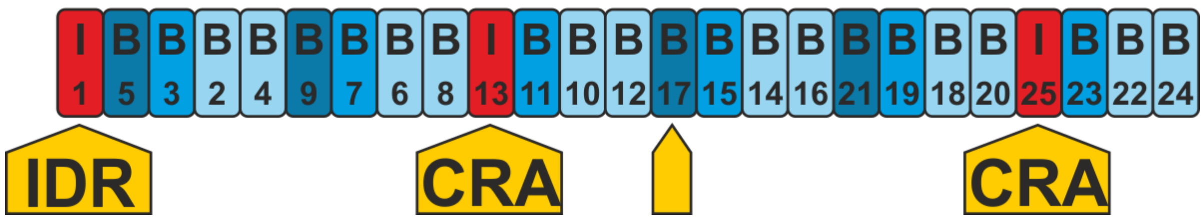
  A coded picture that causes the decoding process to mark all reference pictures except the current CRA picture as "unused for reference" immediately before the decoding of the first picture following the current CRA with an output order greater than the current CRA picture. All coded pictures that follow a CRA picture in output order can be decoded without inter prediction from any picture that precedes the CDR picture in output order. All coded pictures with output order smaller than the current CDR are not affected by the deferred marking process.

Open GOP using CRA pictures

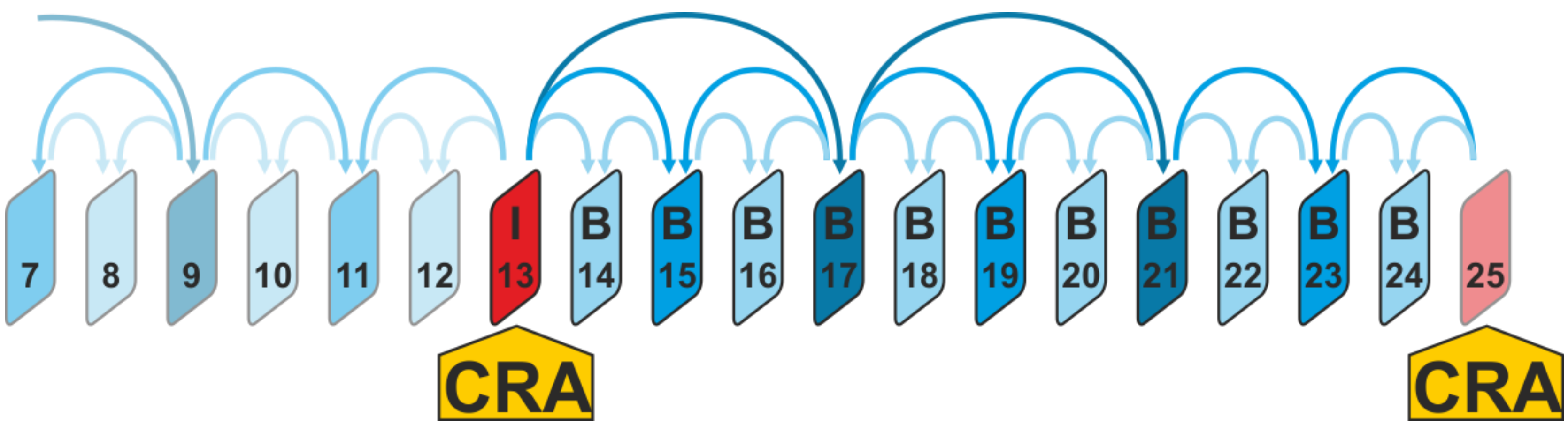
Picture display order:



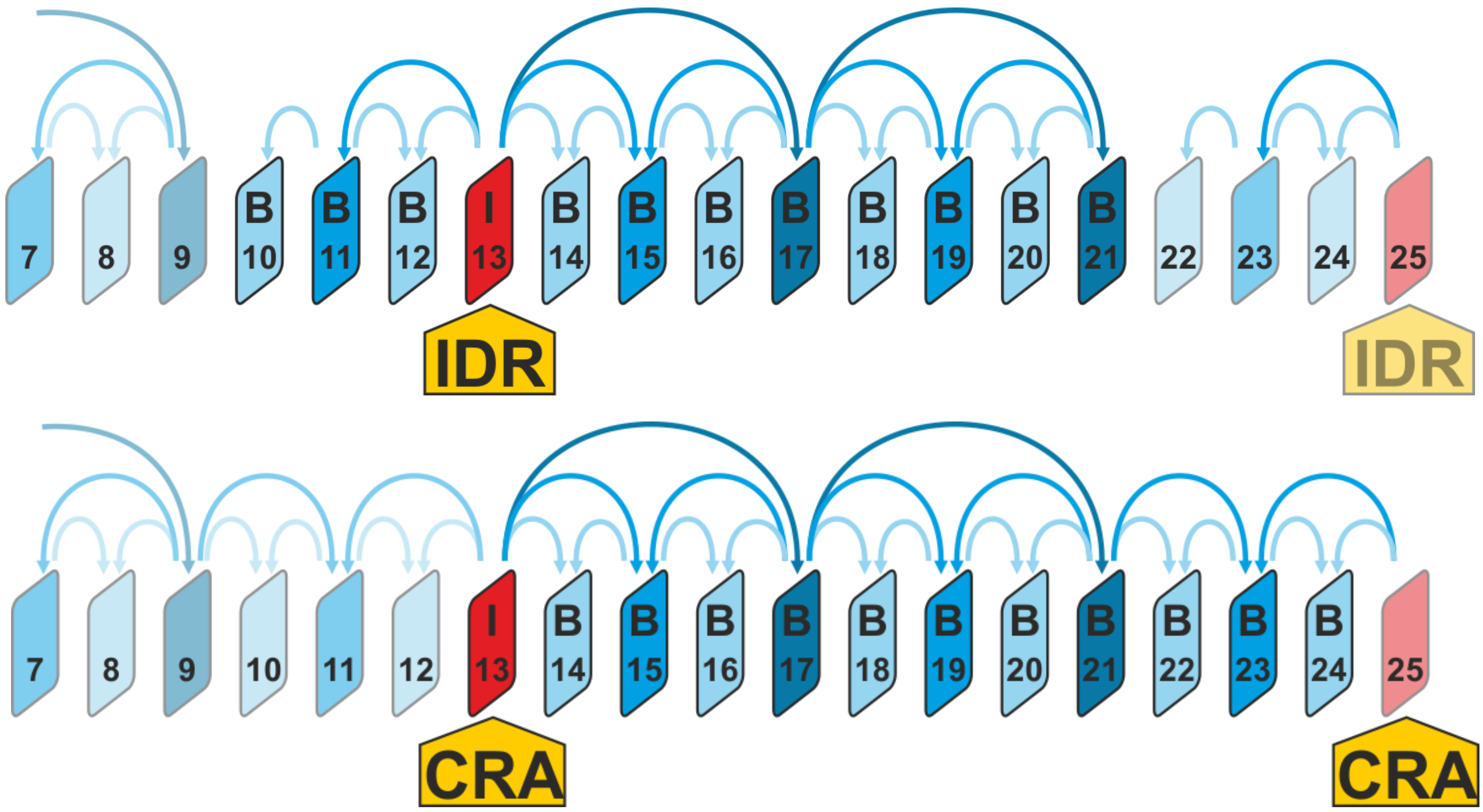
Picture coding order:



GOP shape and possible inter prediction sources:



Comparison: IDR vs CRA



- Coding efficiency improvement (~6%)

Parallelization techniques

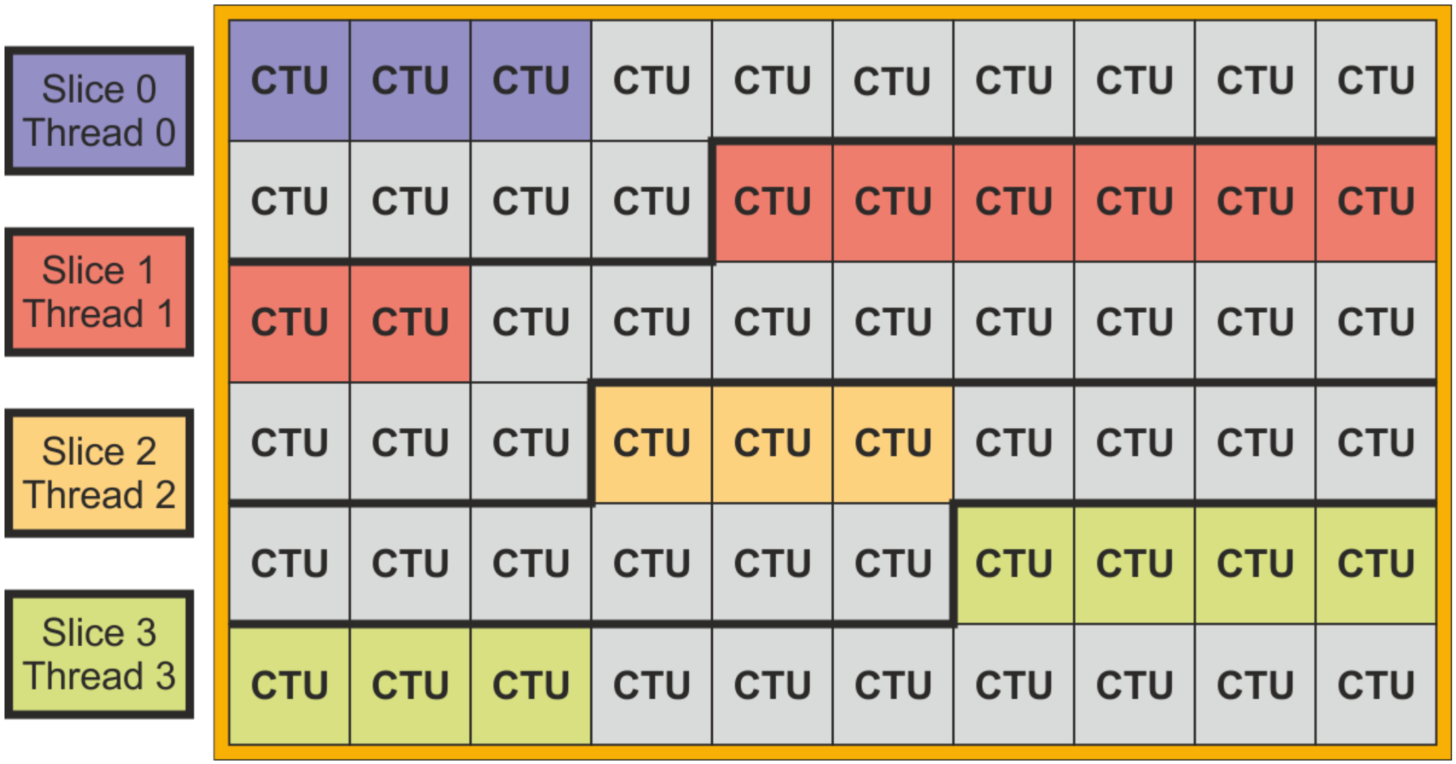
- Most modern multimedia devices have multi- or many- core CPUs or low frequency parallel accelerators
- Video compresion process is difficult to parallelize
- Efficient parallel encoding/decoding reqiures normative techniques

- HEVC standard defines three parallelization techniques:
  - Slices
  - Tiles
  - Wavefront

- Each of them is suitable for different coding scenario
- Parallelization breaks dependancy chains
  – compression performance reduction

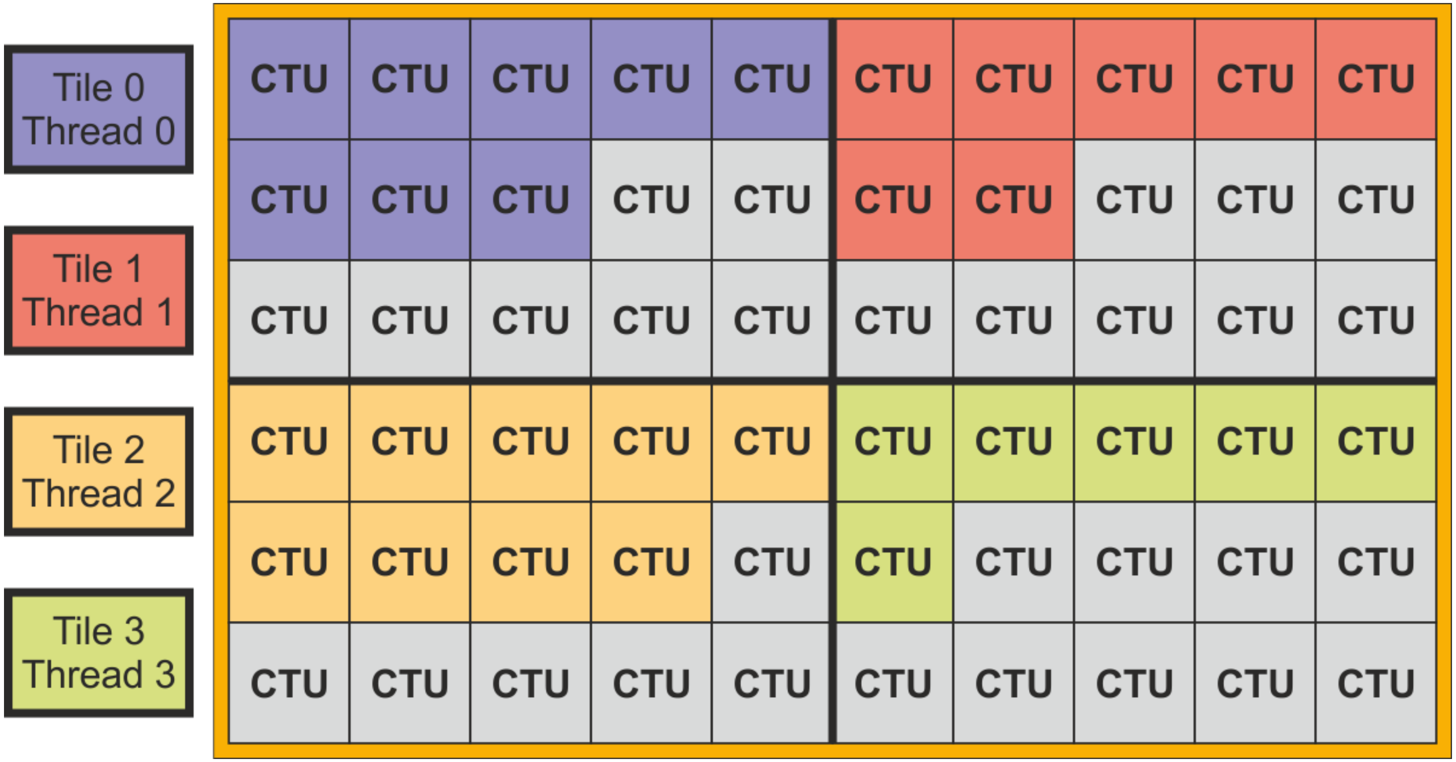
Parallelization: Slices

- Slice boundary breaks any dependencies
  no prediction across slice boundary
- Each Slice is transmitted in separate NALU
- High performance degradation
- Easy to implement



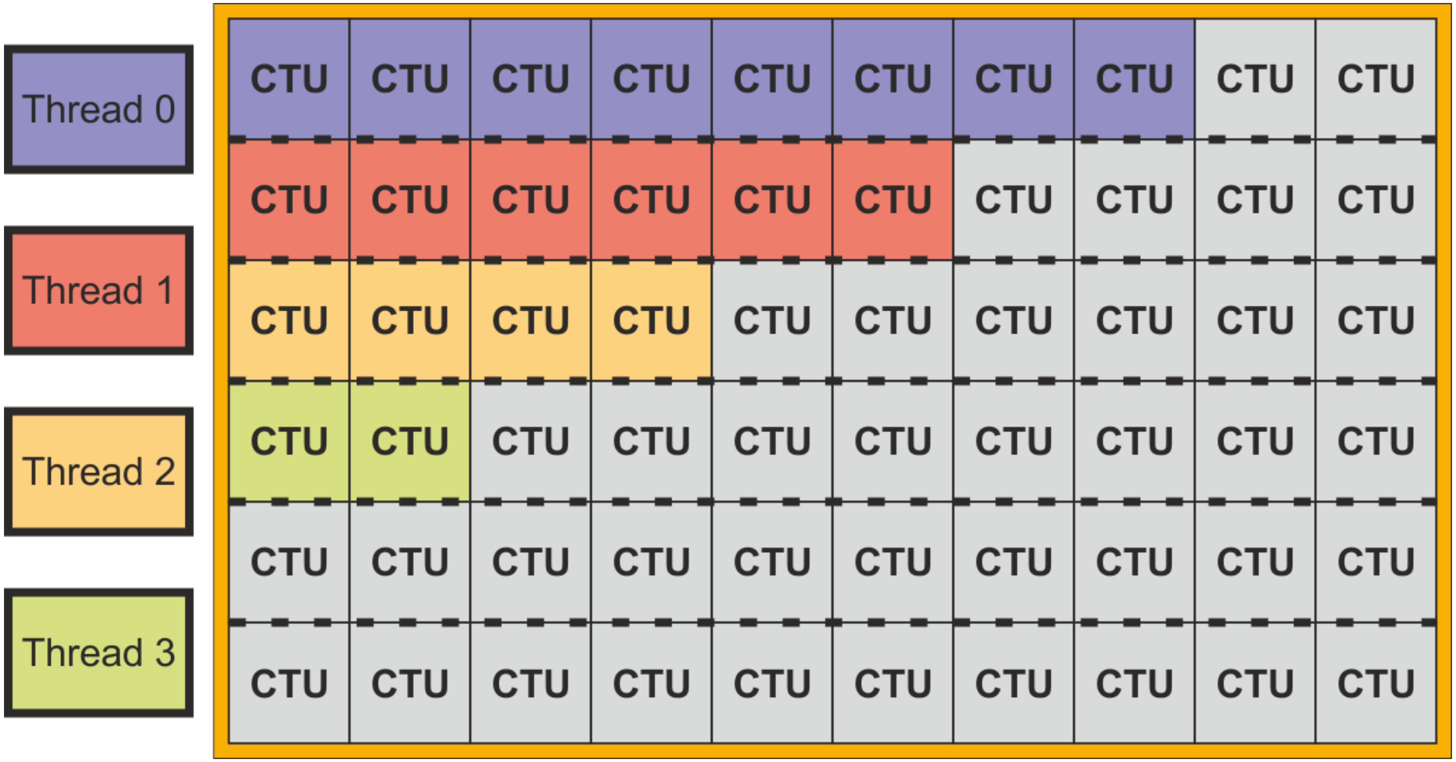
Parallelization: Tiles

- Tile boundary breaks any dependencies
  no prediction across tile boundary
- All tiles can be transmitten in one Slice (and one NALU)
- Suitable for sensors witch multiple readout chanels
- High performance degradation



Parallelization: Wavefronts

- Each CTU row encoded by separate thread
- Prediction cross row boundary
- Very low performance degradation
- Prolog and epilog effect
- Suitable for general usage and ultra-low latency systems



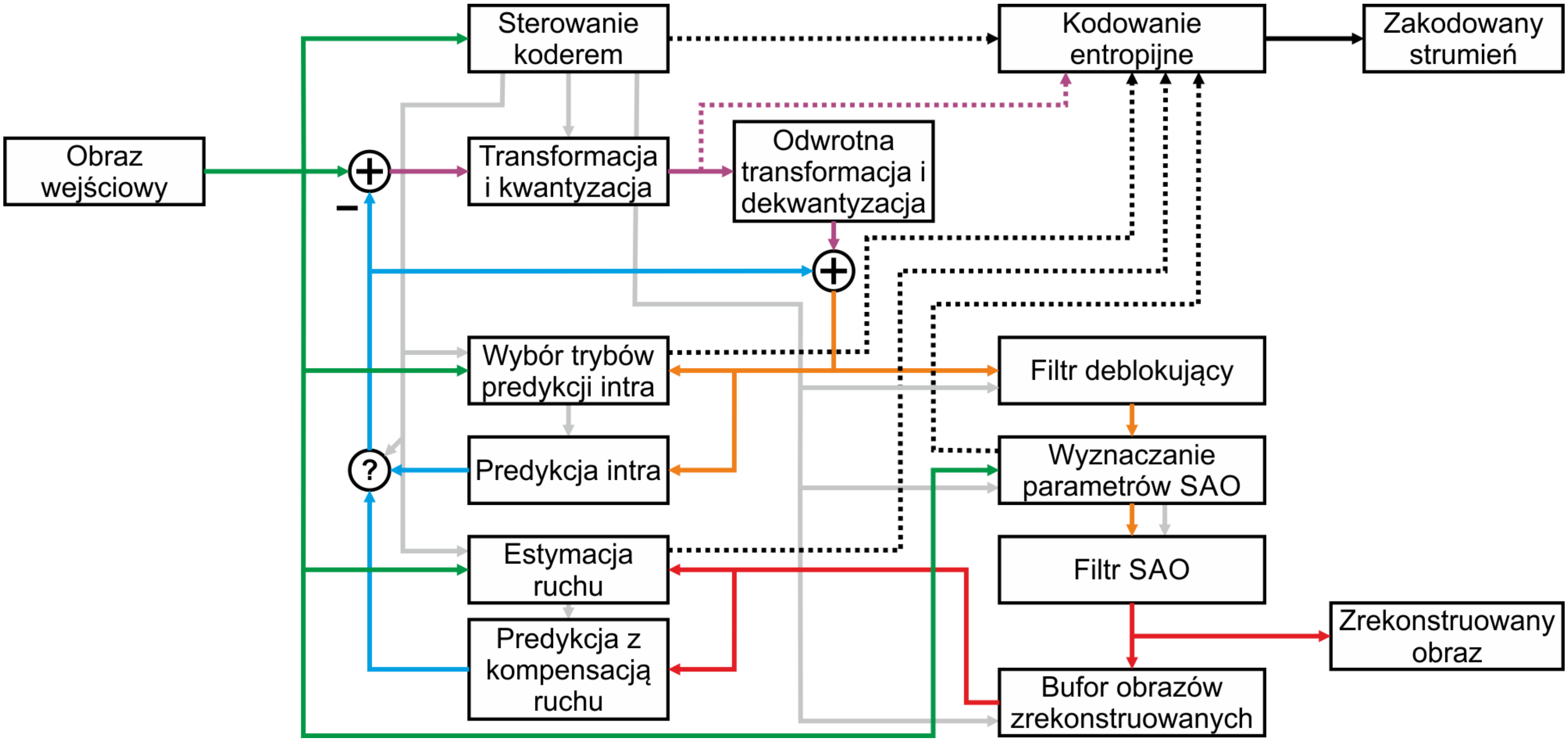
Parallelization techniques combining

- Some of parallelization techniques can be combined together

- Slices + Tiles
  - Multiple tiles within one Slice (in one NAL Unit)
  - Every Tile within one Slice
  - Multiple slices within one Tile

- Slices + Wavefront
  - Extream case: every CTU row in separated dependent slice (ultra-low latency and medicorate performance degradation)

Encoder block diagram

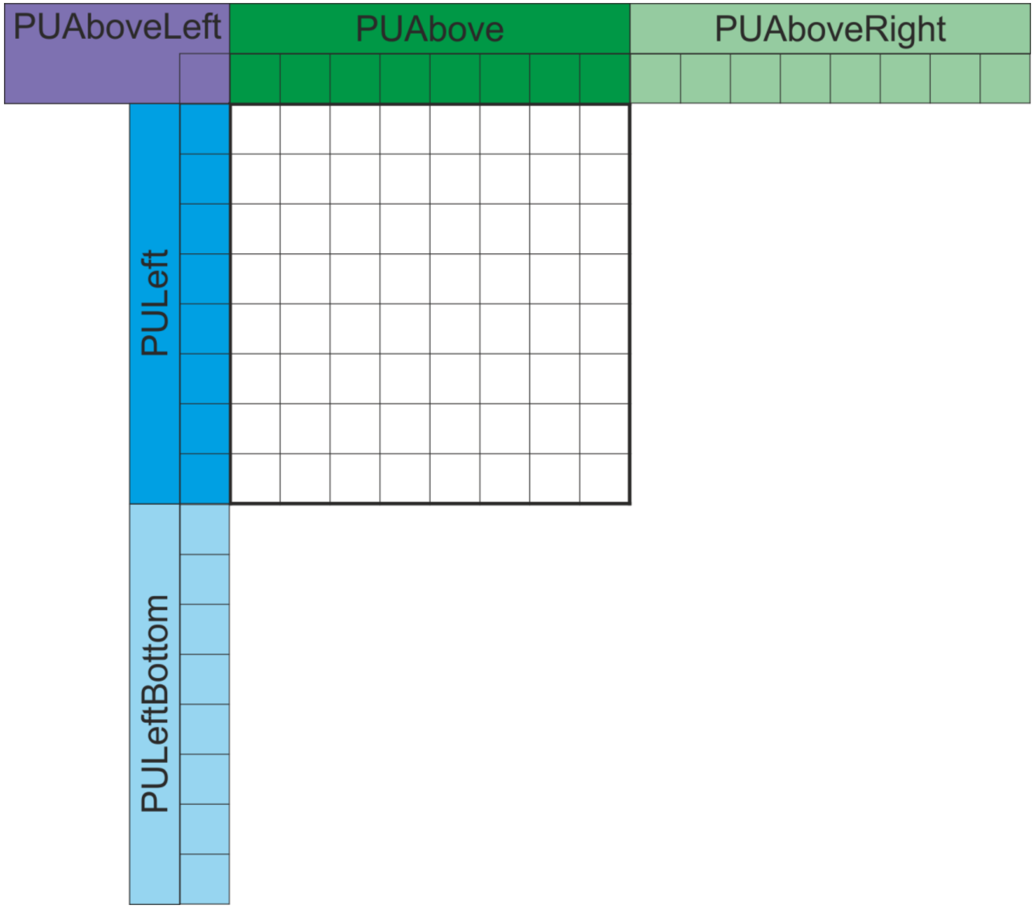


Intra prediction

- Uses encoded/decoded blocks boundary samples to generate prediction signal
- Exploits local image correlation and directional characteristics
- Available in I-, P- and B-slices
- Improved compared to AVC
- 35 prediction modes available for all PU sizes
(in AVC 9 modes for 4x4/8x8 and 4 modes for 16x16)
- Efective prediction mode encoding
  - 3 MPM (most probable modes)
  - remaining modes (32) with MPM ellimination

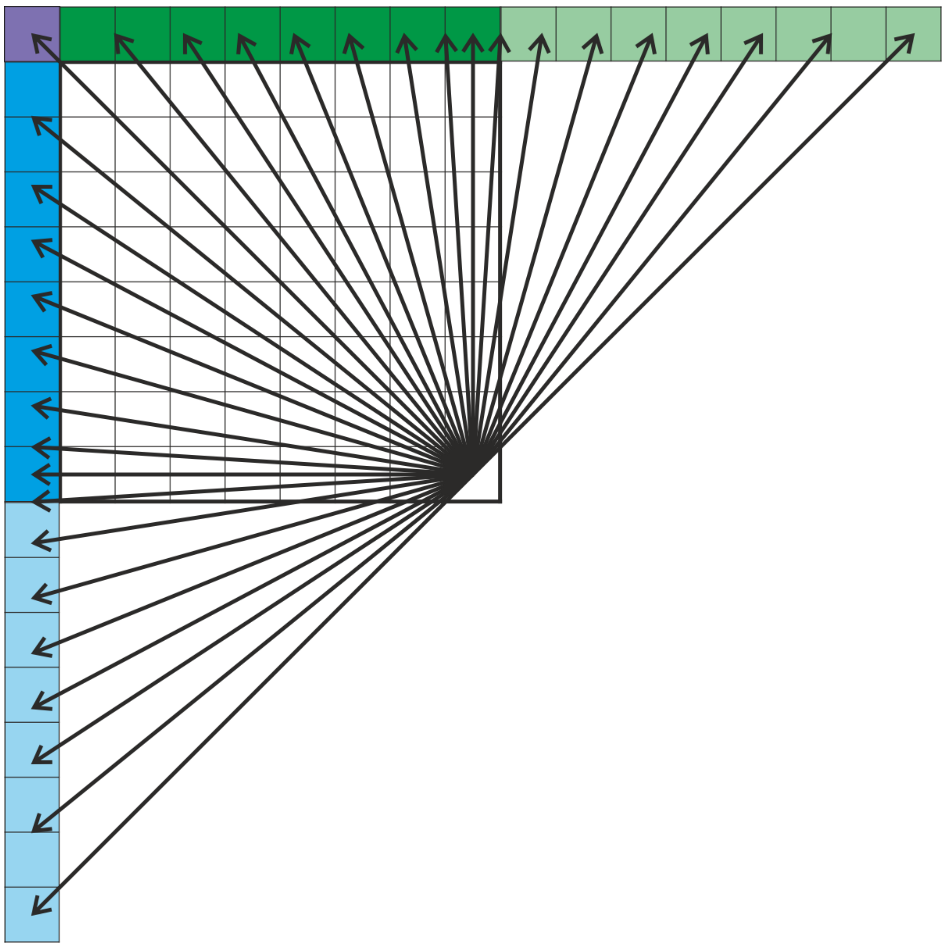
Reference samples

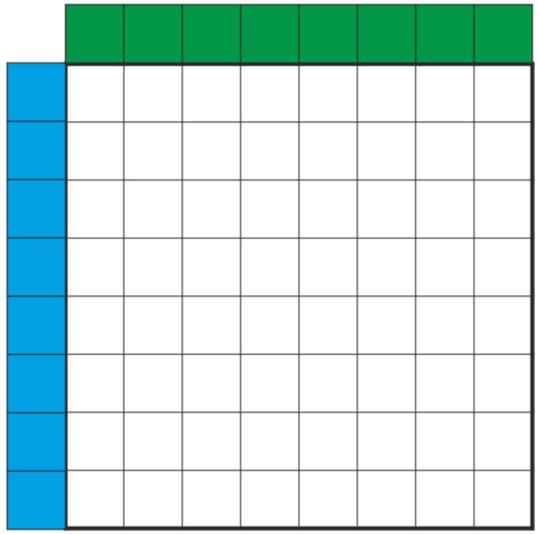
- Reference samples are derived from neighboring blocks
- Not available samples are substituted
- Reference samples filtering
  - dependent on PU size and prediction mode
  - low pass FIR filter (1 2 1)/4

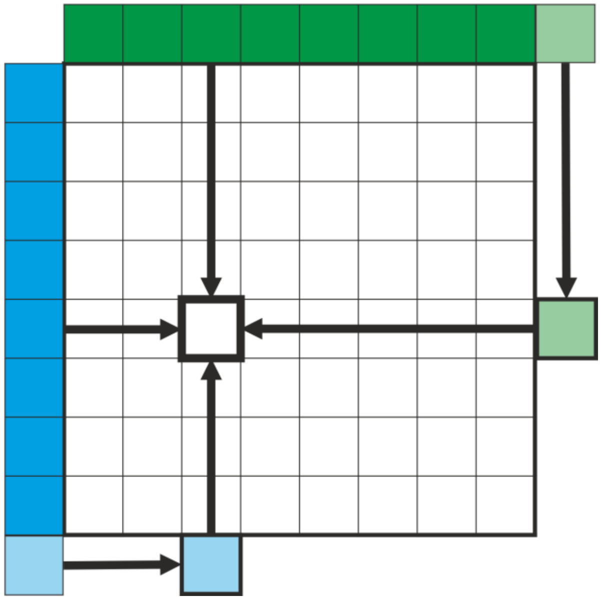


Prediction modes

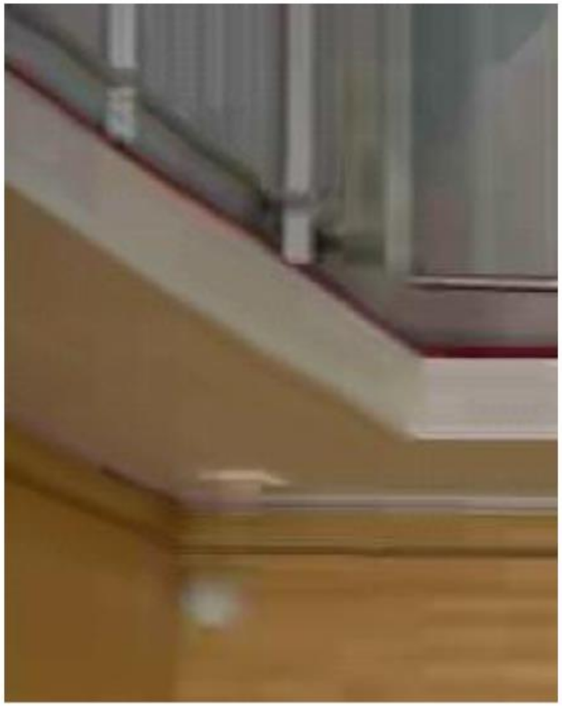
- Mode 0: DC (Average value)
- Mode 1: Planar (Bidirectional weighted)
- Mode 2-34: Angular



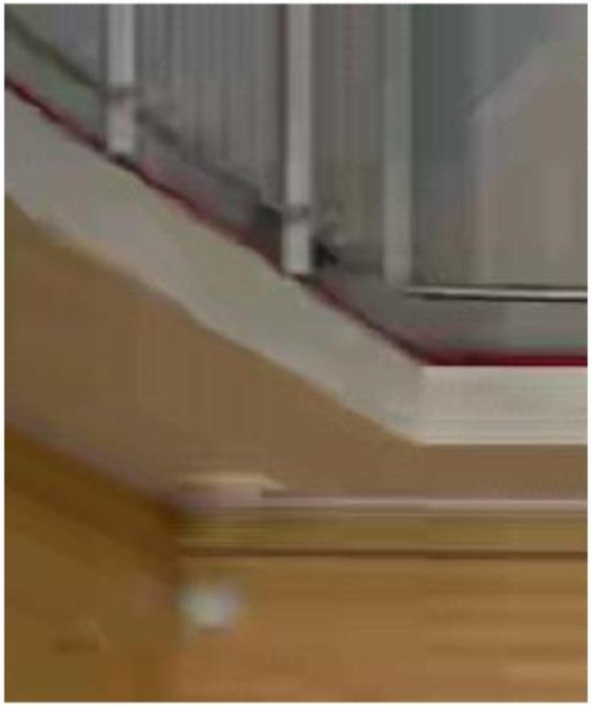




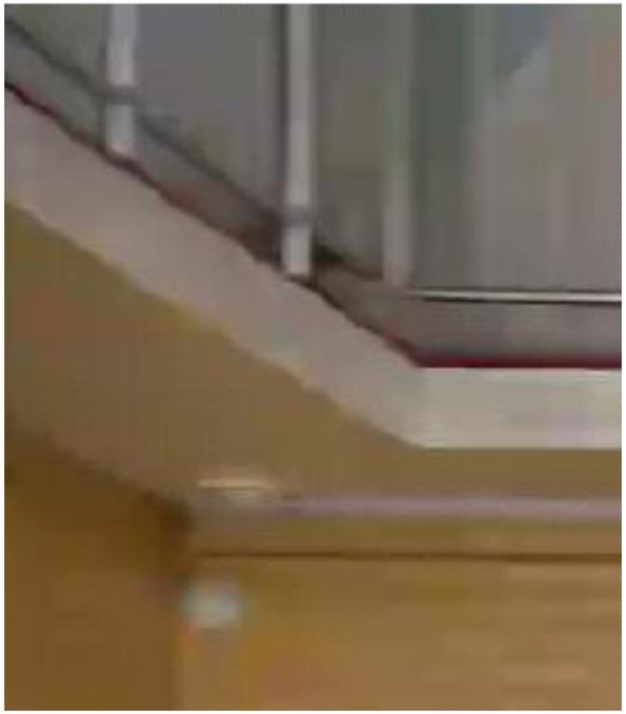
Intra prediction comparison



HEVC encoder



HEVC encoder using AVC prediction modes



AVC encoder

Inter (motion compensated) prediction

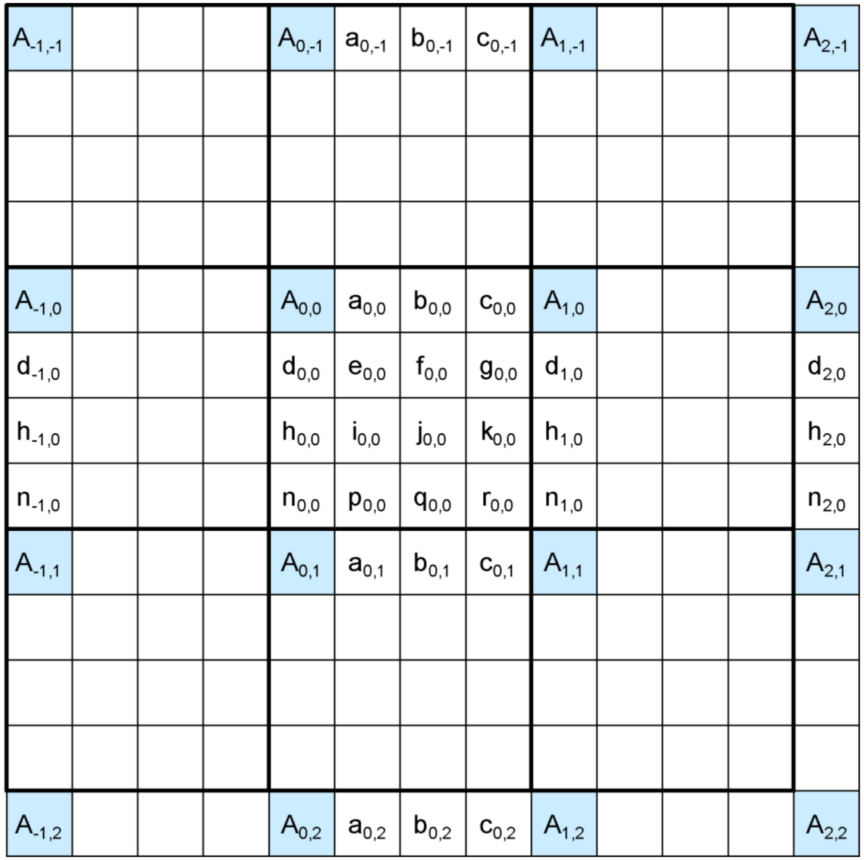
- Uses previously encoded/decoded pictures to generate prediction signal
- Uni- and Bi-directional prediction
- Motion vectors with ¼ pel precision
- Reference image interpolation (½, ¼ and ¾ position)
- Improved interpolation filter
- Reduced rounding error
- Improved motion vector prediction

Interpolation filter

- Integer samples (A)
- Half pel samples (b, h, j)
- Quarter pel samples (a, c, d, e, g, i, k, n, p, q, r)

- Interpolation filter derrived from DCT transform

- Half pel filter coefficients: (-1 4 -11 40 40 -11 4 -1)
- Quarter pel filter coefficients:
  (-1 4 -10 58 17 -5 1 0)
  (0 1 -5 17 58 -10 4 -1)

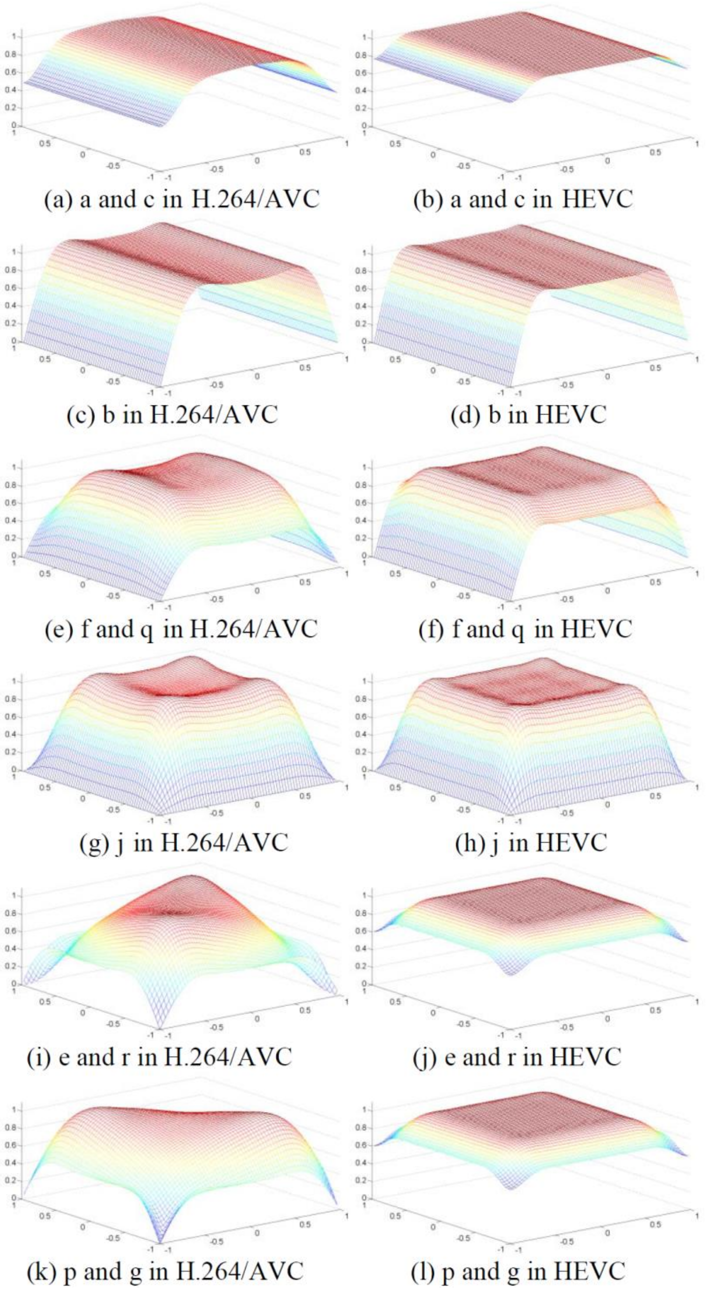


Interpolation filter comparison

- Magnitude response of interpolation filters →
- Complexity:
  (average number of operations needed to get one sample of prediction signal)

	HEVC	AVC
Pixel accesses	33.06	16.31
Multiplications	37.13	18.38
Additions	32.06	16.06

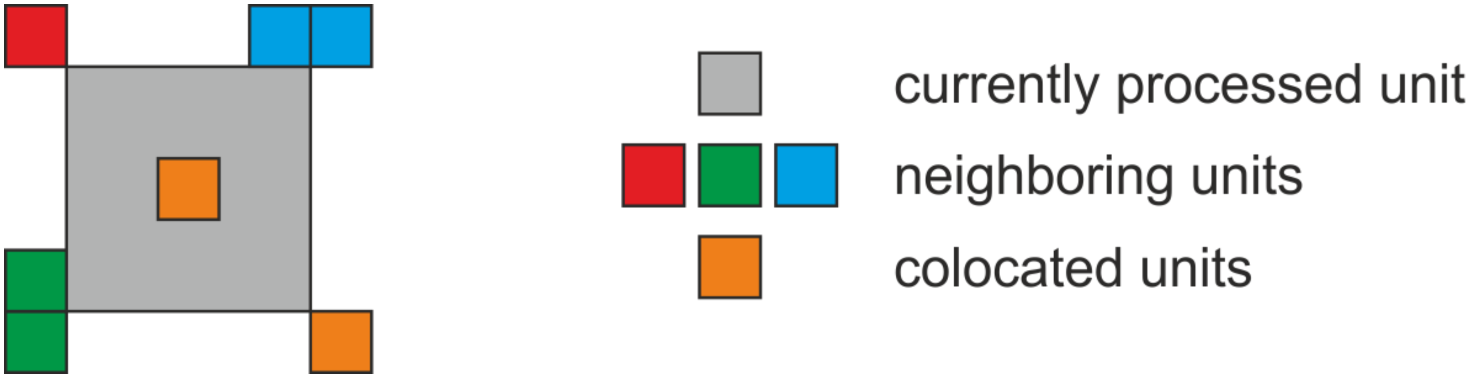
- Performance:
  HEVC with AVC interpolation filter:
  - low delay coding (IPPPPPPP...)
  - average **15%** bitrate increase
  - mostly due to q-pel filter



(a) a and c in H.264/AVC
(b) a and c in HEVC
(c) b in H.264/AVC
(d) b in HEVC
(e) f and q in H.264/AVC
(f) f and q in HEVC
(g) j in H.264/AVC
(h) j in HEVC
(i) e and r in H.264/AVC
(j) e and r in HEVC
(k) p and g in H.264/AVC
(l) p and g in HEVC

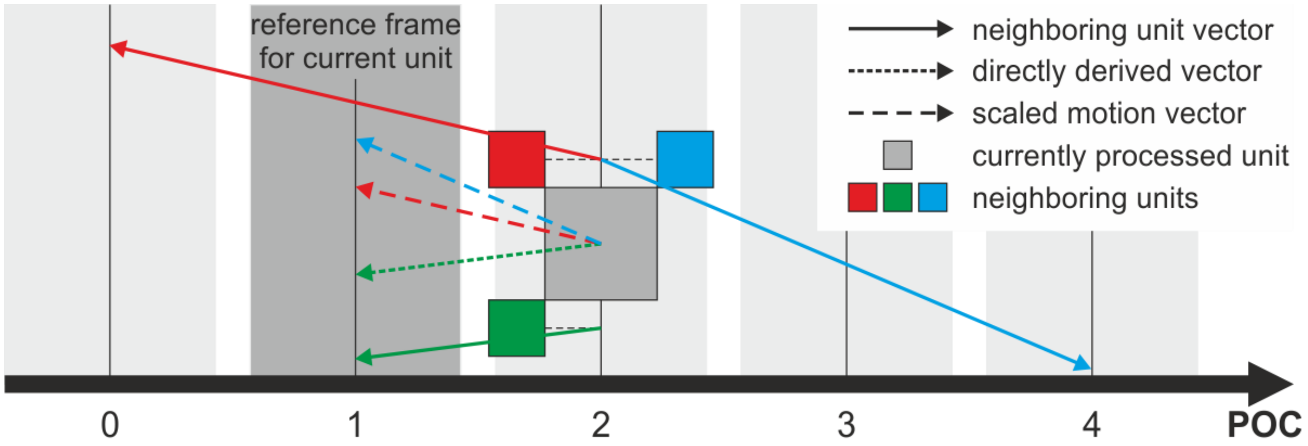
Motion vector prediction

- Motion vector prediction in HEVC:
  AMVP (Advanced Motion Vector Prediction)
- Two motion vector prediction modes:
  - Motion vector competition
  - Motion data merging
- Prediction from previously encoded frames:
  TMVP (Temporal Motion Vector Prediction)
- Prediction candidates:
  - spatial (naighboring PUs)
  - temporal (colocated PU in reference frame)

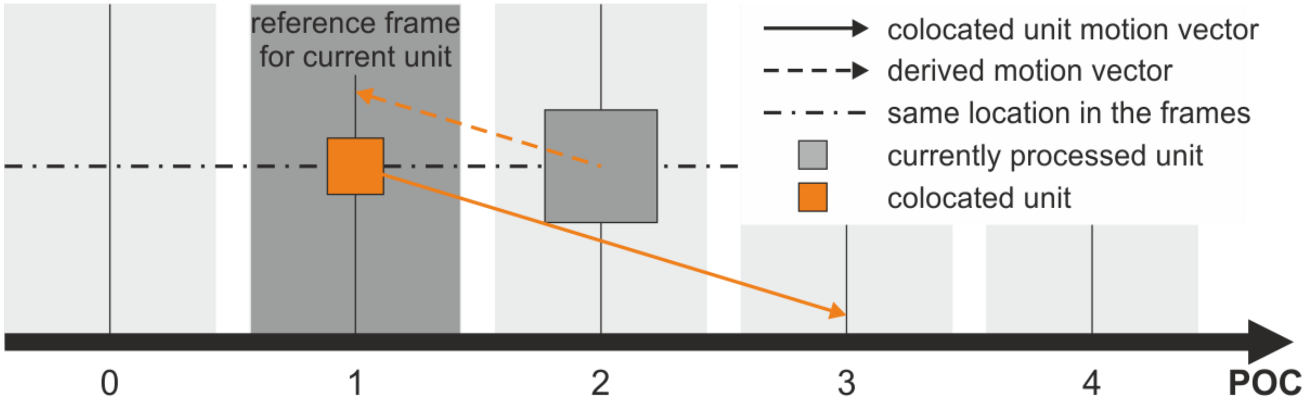


Motion vector scaling & colocated unit

- Motion vector scaling



- Prediction from colocated PU

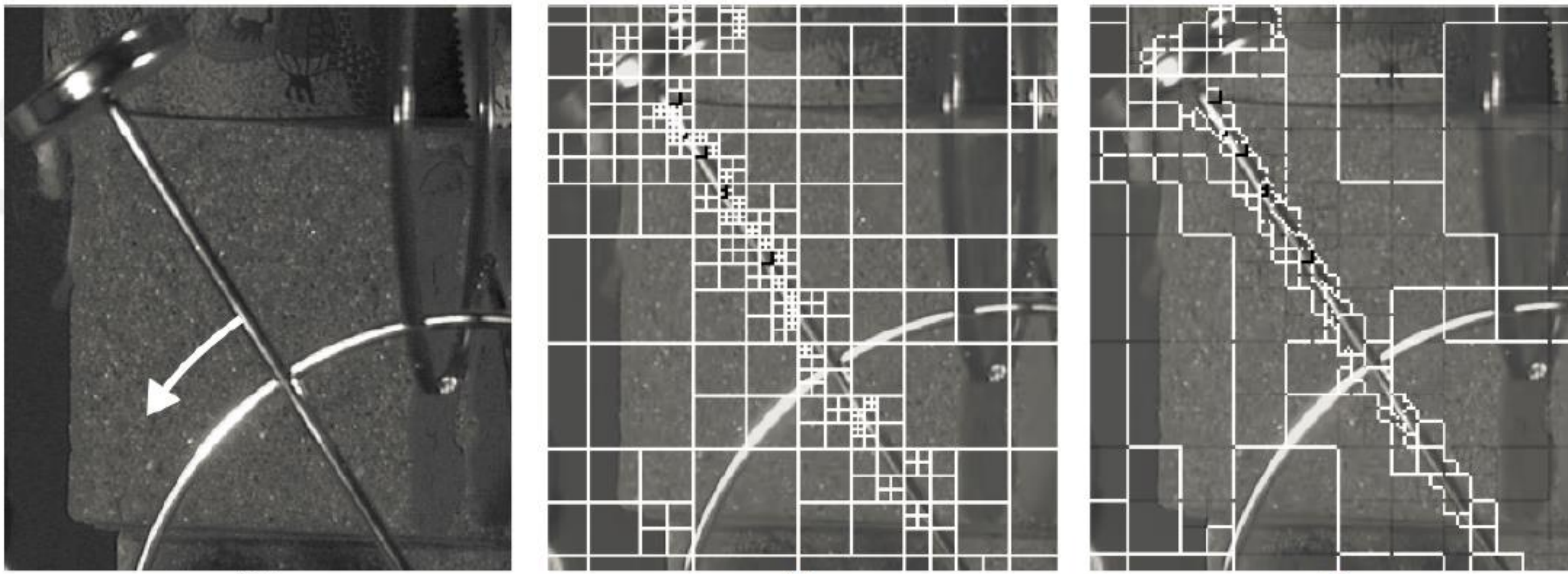


AMVP

- Motion vector competition:
  - Motion vector is predicted using 2 candidates (2 spatial or 1 spatial + 1 temporal)
  - Encoder transmits:
    - Prediction mode (uni or bi)
    - Reference picture identifier(s)
    - Motion vector prediction identifier(s)
    - Motion vector residual(s)

- Merge mode:
  - Motion field (prediction mode, motion vector(s), reference picture identifier(s)) are derrived from spatial or temporal candidate
  - 5 Merge candidates (up to 4 spatial and up to 2 temporal)
  - Encoder transmits:
    - Merge candidate identifier
  - Merge mode is used for Skip CUs

Merge mode



(a) Moving Object

(b) Without Merge (many extra motion parameters)

(c) With Merge

Residual coding

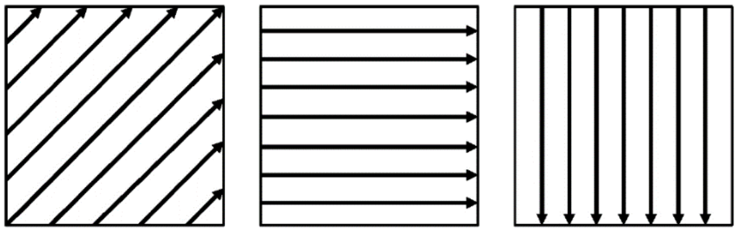
- Residual signal is transformed, quantized, scanned and entropy encoded

- Transforms:
  - DCT (Discrete Cosine Transform) – 4x4, 8x8, 16x16 and 32x32 size
  - DST (Discrete Sine Transform) – Intra 4x4 only
  - TS (Transform Skip) mode – 4x4 only **(no transformation !!!)**

- Quantization:
  - Introduces compression loss
  - QP (Quantization parameter): 0-51
    - 0 – smallest quality loss, highest bitrate
    - 51 – highest quality loss, smallest bitrate

Residual coding

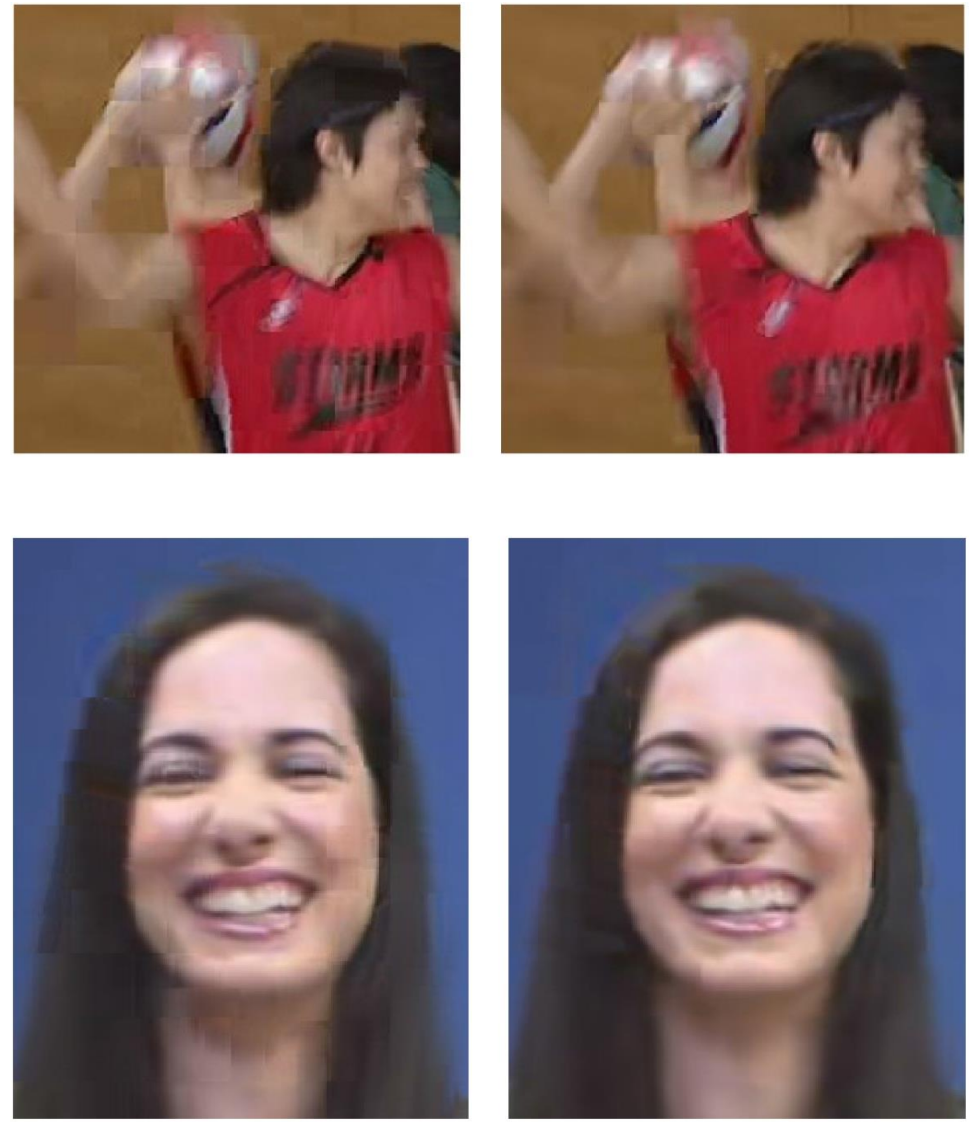
- Scanning:
  - MDCS (Mode Dependent Coefficient Scan) for 4x4 and 8x8 intra
    - Scan type depends on intra prediction mode
  - Diagonal scan
    - all inter predicted CUs,
    - intra DC, Planar, ~45º (diagonal) angular directions
  - Horizontal scan
    - intra vertical (and close to vertical)
  - Vertical scan
    - intra horizontal (and close to horizontal)



- Entropy encoding:
  - CABAC (Context Adaptive Binary Arithmetic Encoder) only
  - Reduced complexity when compared to AVC
  - Improved fine granularity parallelism (hardware decoding)

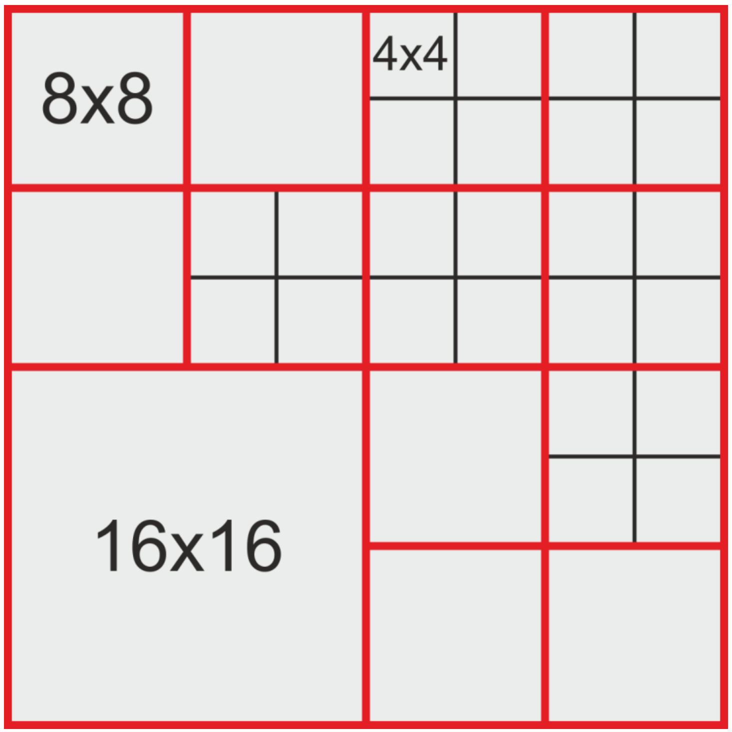
Loop filter - Deblocking

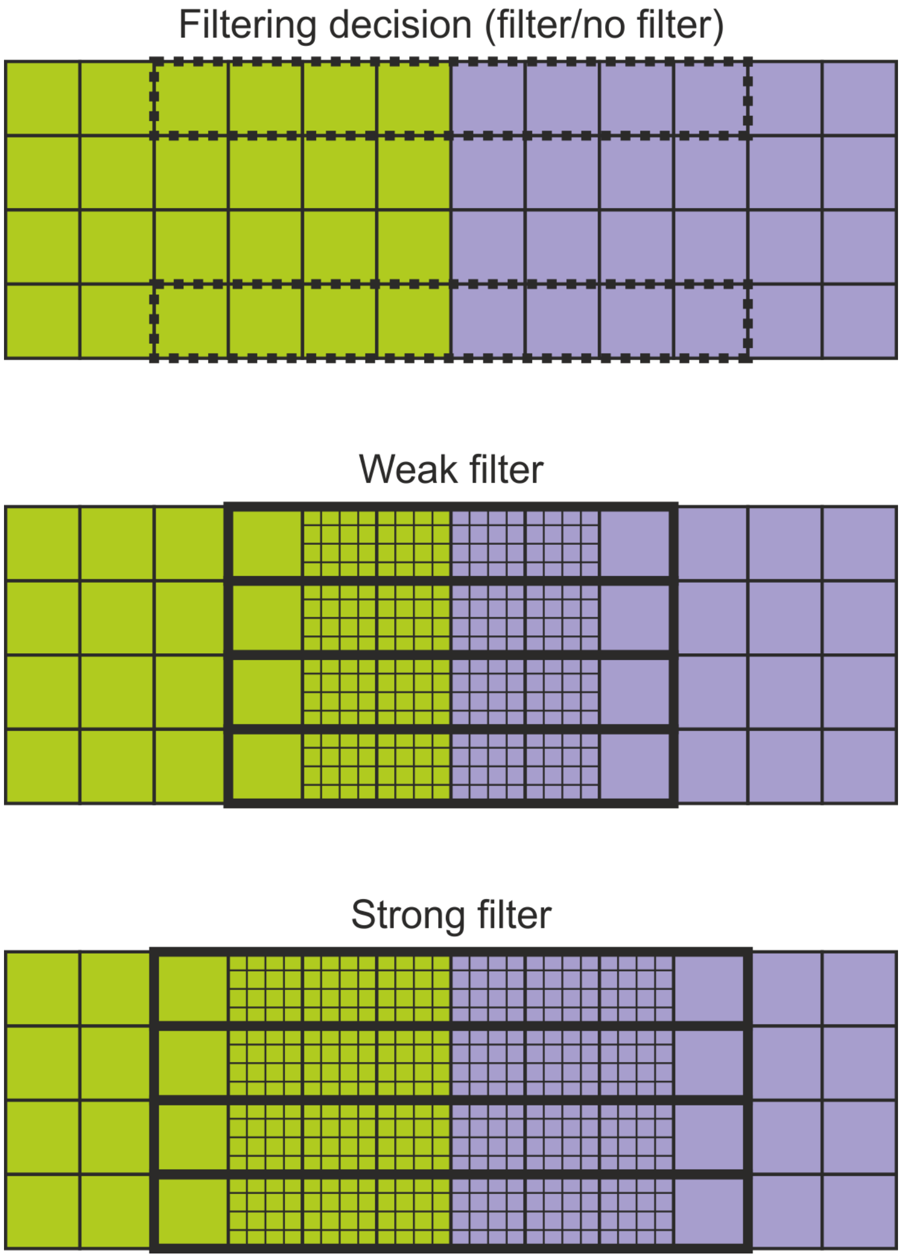
- Reduces **blocking** artifacts
- Aplied to samples at TU and PU boundaries
- 3 filtering types:
  - no filter
  - week filter
  - strong filter
- Filtering type decision:
  - prediction type
  - residual signal presence
  - reference frames
  - motion vectors
  - samples at PU/TU boundary
- Reduced complexity



Loop filter - Deblocking

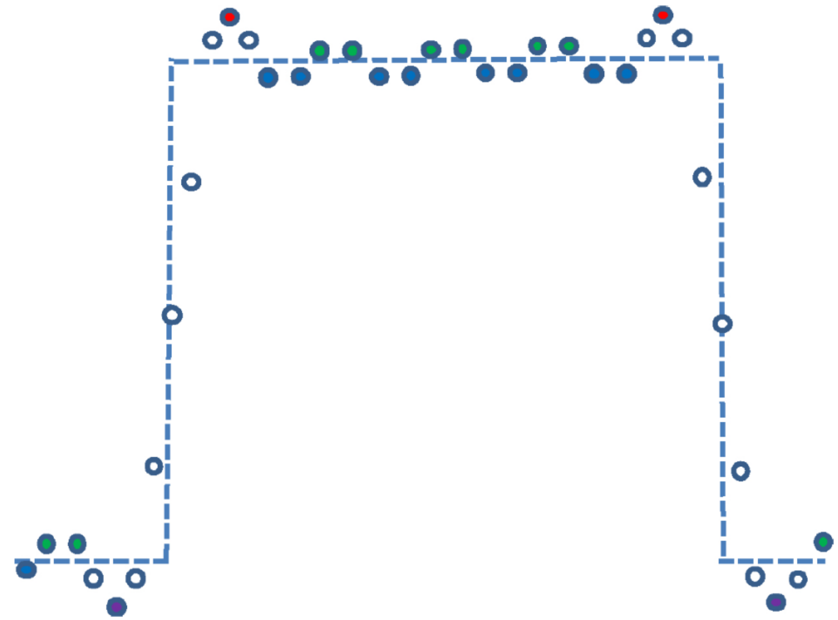
- HEVC only applies the deblocking filter to the edges that are aligned on an 8×8 sample grid
- Reason:
  parallel deblocking

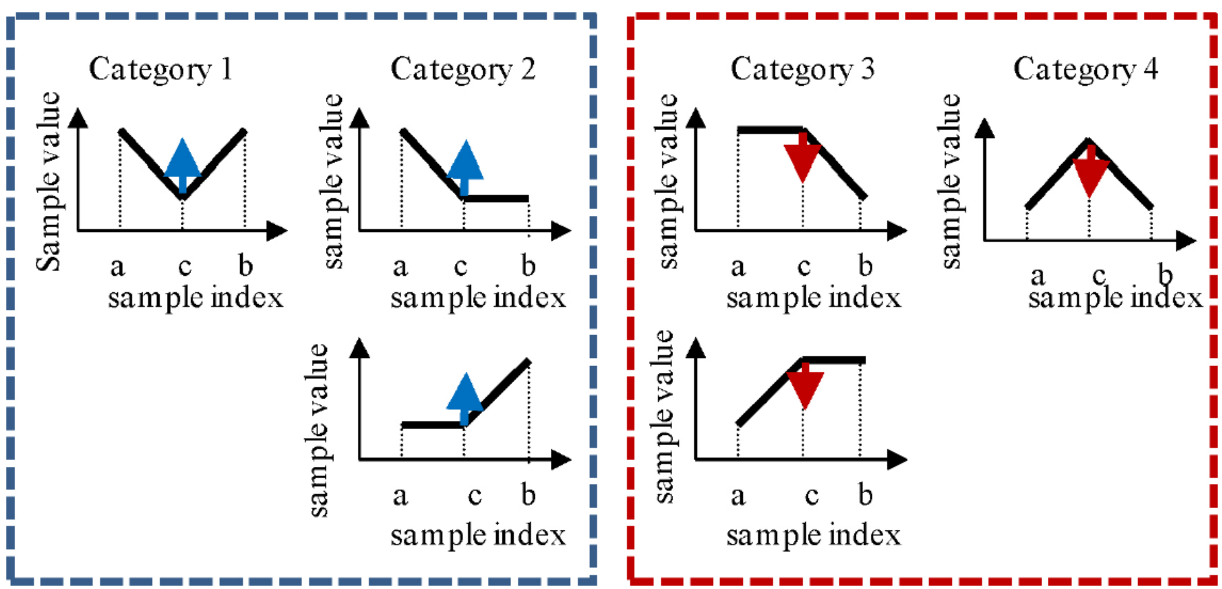




Loop filter - SAO

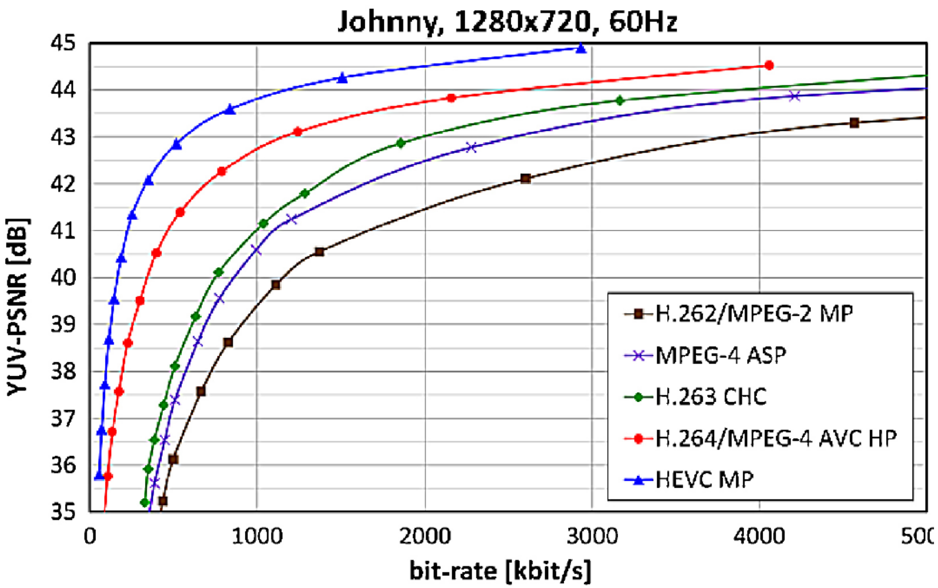
- SAO (Sample Adaptive Offset)
- Reduces **ringing artifacts** caused by transform coefficient quantization (ringing artifacts presence is increased by introduction of large transforms)
- Operates as an additional in-loop filter
- Two offset types:
  - Edge Offset (EO)
  - Band Offset (BO)

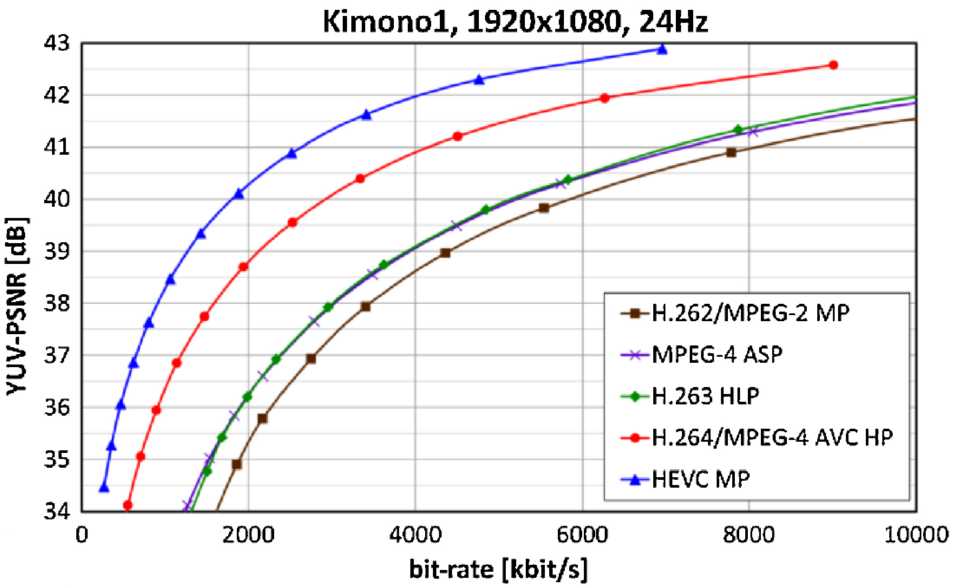


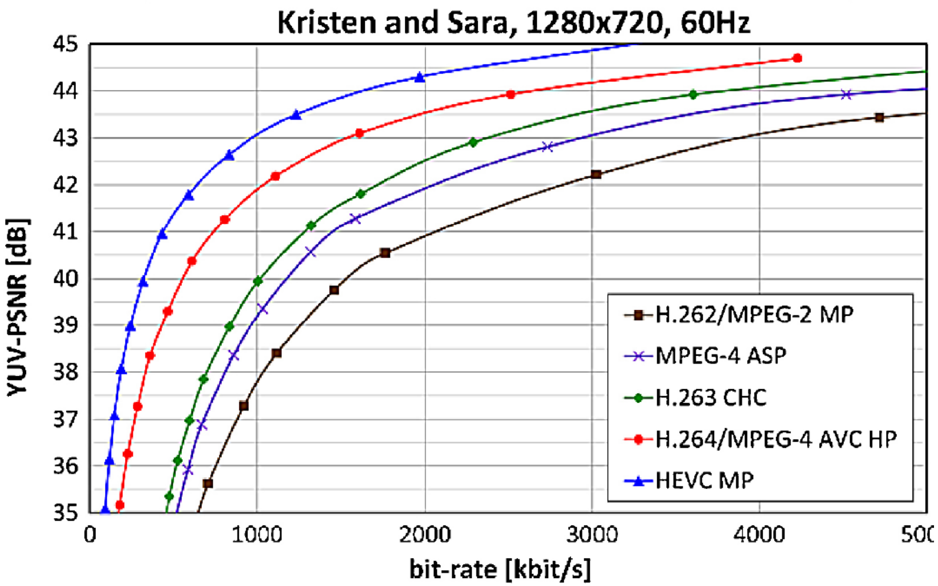


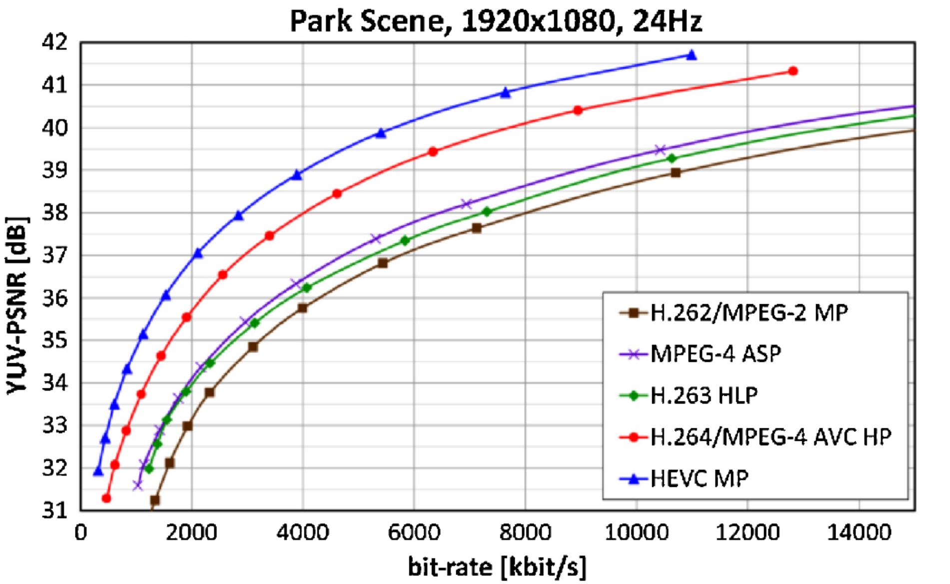
Loop filter – SAO – example

Objective performance comparison

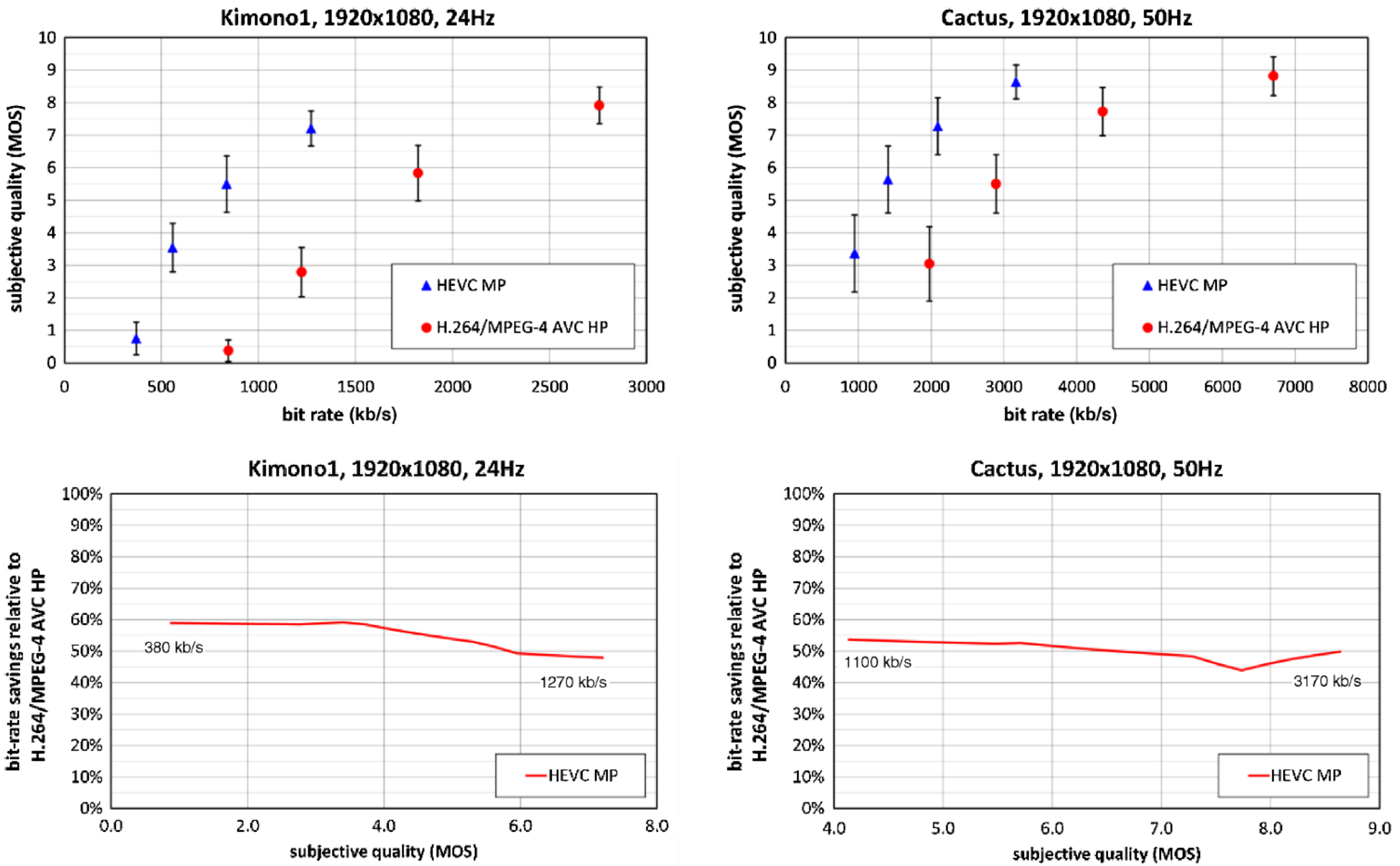








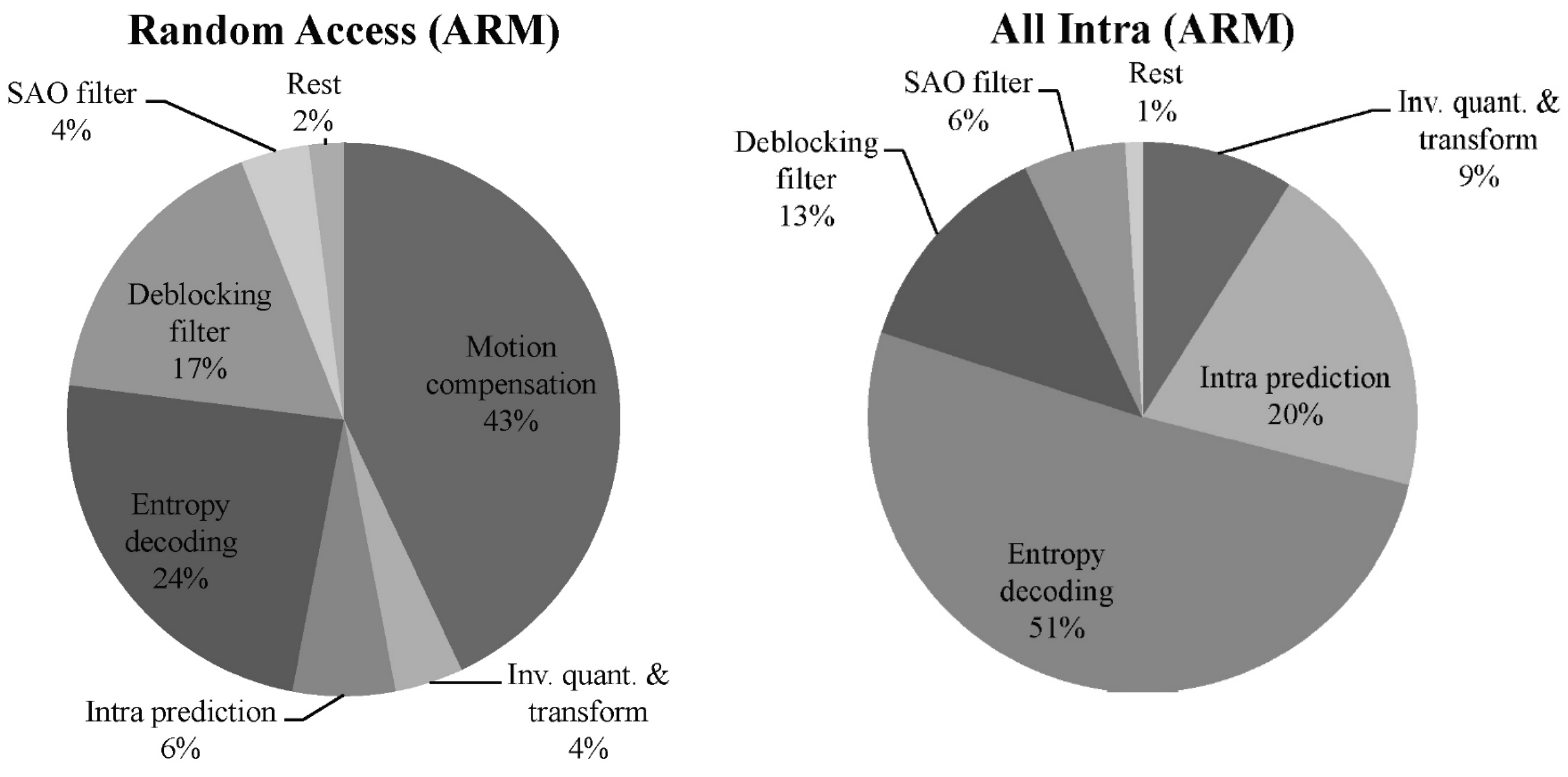
Subjective performance comparison



Decoder performance

- **Optimized HEVC decoder has performance similar to optimizer AVC decoder !!!**

- Profiling of the NEON optimized HEVC decoder:

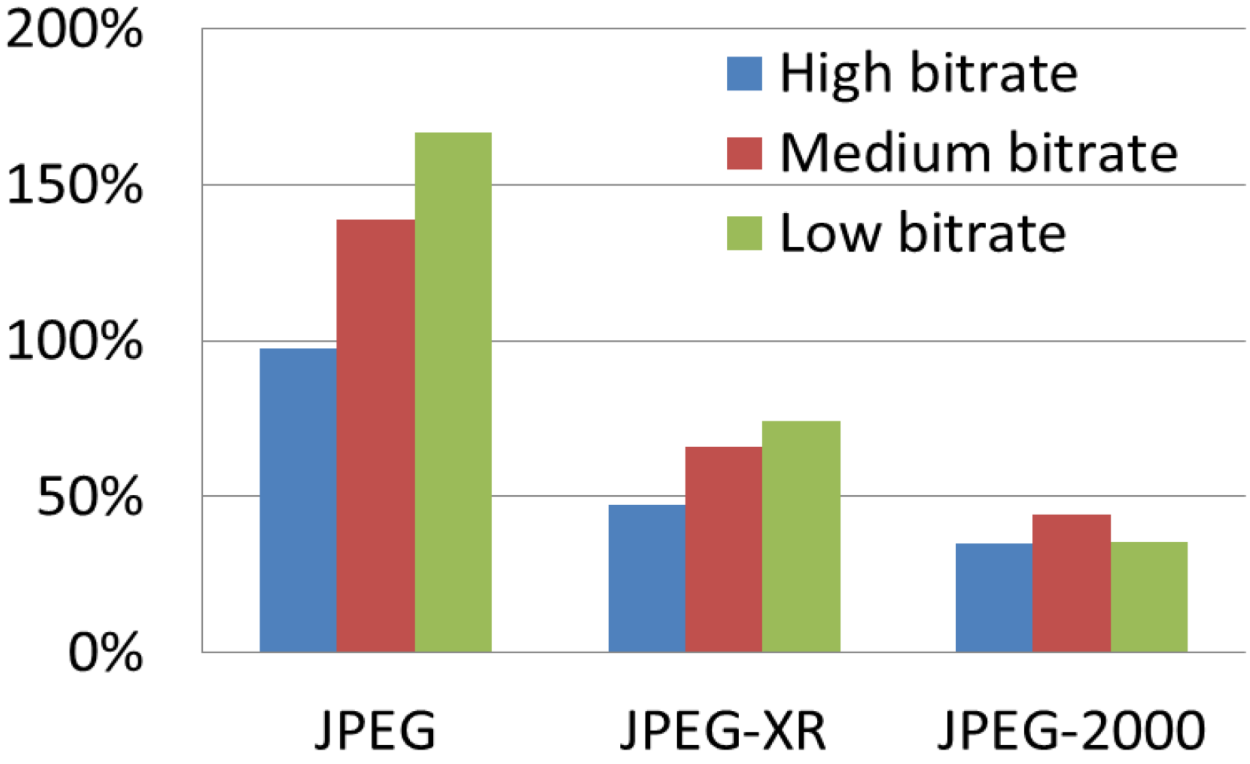


Still Picture Profile

- Profile dedicated for transmission of single picture
- Coding efficiecy higher than JPEG 2000 and JPEG XR

Additional rate required by JPEG to achieve the same quality:

(and 2x times more for computer generated graphics)



Examples

- Compressed image
- Prediction signal
- Residual signal

- CU/PU/TU size selection
- CBFY units