# Recent Developments in Video Compression Standardization

**CVPR CLIC Workshop, Salt Lake City, 2018-06-18**

Jens-Rainer Ohm
Institute of Communication Engineering
RWTH Aachen University
ohm@ient.rwth-aachen.de

## Outline

1. Introduction and history of video coding standardization
2. Call for Proposals on Versatile Video Coding
3. Tools for improved compression
4. Methods related to deep learning

# 1. Introduction and history of video coding standardization

**Recent developments in video compression standardization**

Jens-Rainer Ohm
Institute of Communication Engineering
RWTH Aachen University
ohm@ient.rwth-aachen.de

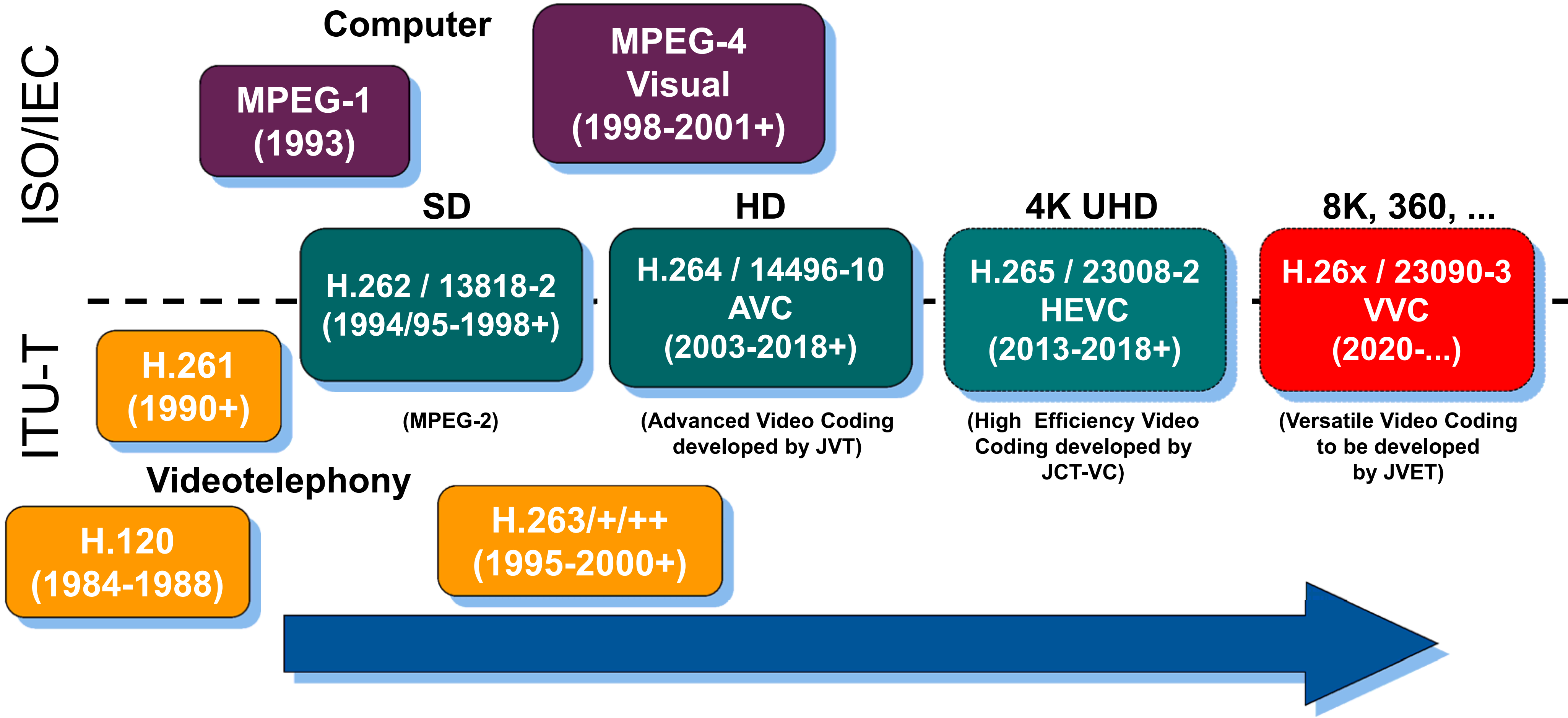
# Motivation for permanent improvements in video compression

- **Video continually increasing by resolution**
  - HD, UHD (4Kx2K, 8Kx4K) appearing
  - Mobile services going towards HD/UHD
  - Stereo, multi-view, 360° video
- **Video has multiple dimensions to grow the data rate**
  - Frame resolution, Temporal resolution
  - Color resolution, bit depth
  - Multi-view
  - Visible distortion still an issue with existing networks
- **Necessary video data rate still grows faster than feasible network transport capacities**
  - Better video compression (50% rate of current HEVC) needed, even after availability of 5G
- **Machine/computer vision applications are also hungry for more video data**
  - For these, stability of feature recognition is probably more important than subjective quality

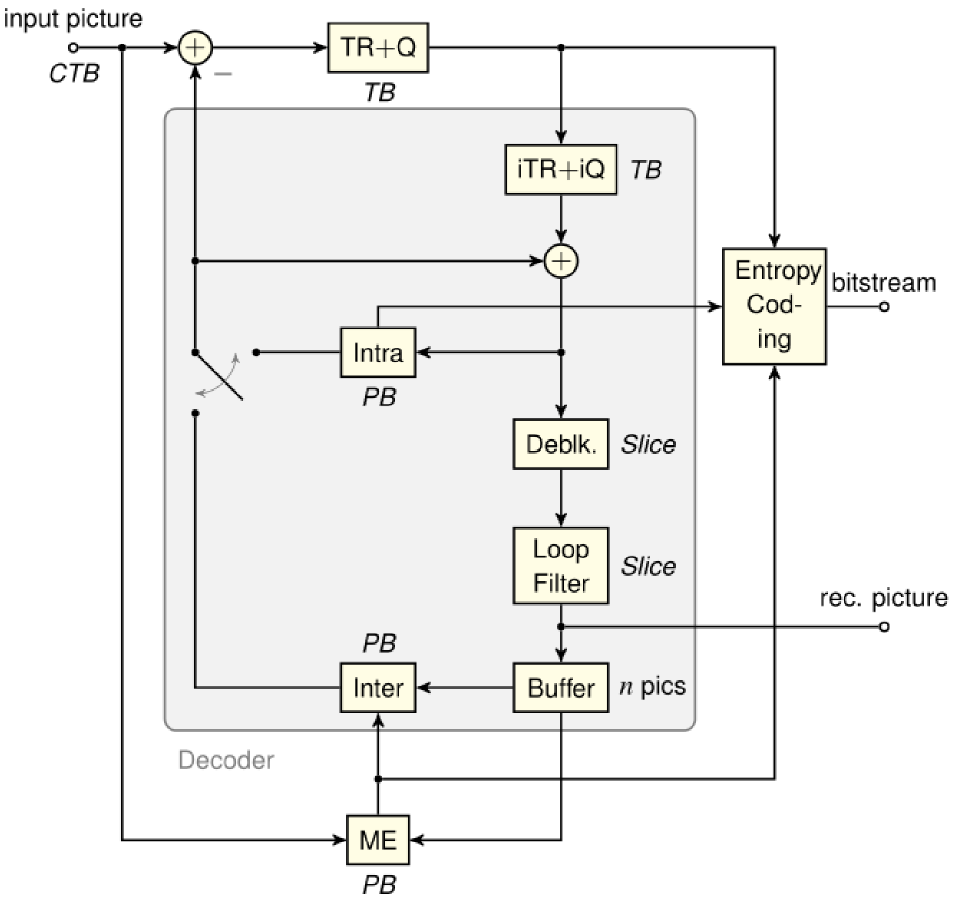
## Video coding standardization organisations

- **ISO/IEC MPEG** = "**Moving Picture Experts Group**"
  (ISO/IEC JTC 1/SC 29/WG 11 = International Standardization Organization and International Electrotechnical Commission, Joint Technical Committee 1, Subcommittee 29, Working Group 11)
- **ITU-T VCEG** = "**Video Coding Experts Group**"
  (ITU-T SG16/Q6 = International Telecommunications Union – Telecommunications Standardization Sector (ITU-T, a United Nations Organization, formerly CCITT),
  Study Group 16, Working Party 3, Question 6)
- **JVT** = "**Joint Video Team**" collaborative team of MPEG & VCEG, responsible for developing AVC (discontinued in 2009)
- **JCT-VC** = "**Joint Collaborative Team on Video Coding**" team of MPEG & VCEG , responsible for developing HEVC (established January 2010)
- **JVET** = "**Joint Video Exploration Team**" exploring potential for new technology beyond HEVC (established Oct. 2015) – renamed to "**Joint Video Experts Team**" responsible for developing VVC from April 2018

# History of international video coding standardization (1985 ~ 2020)



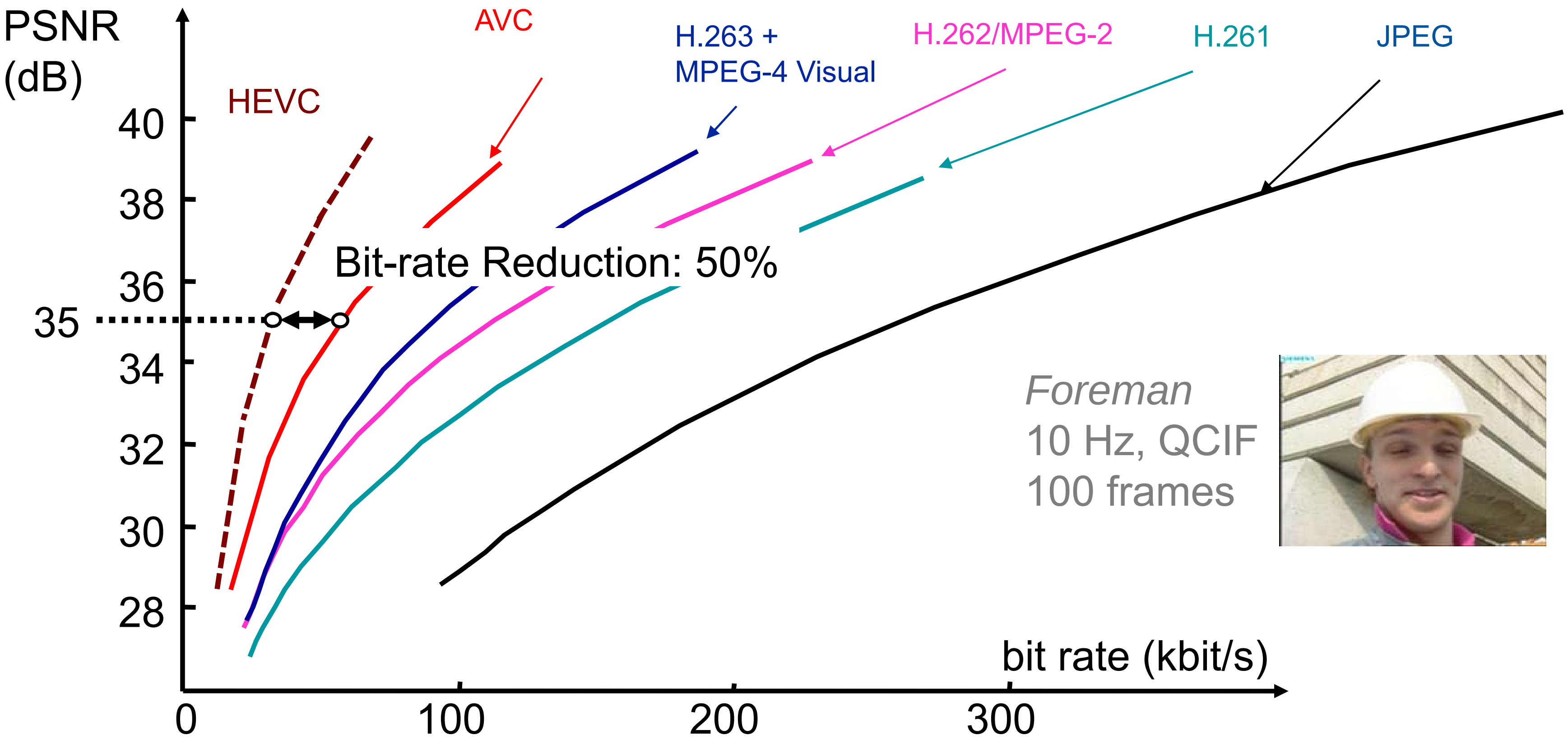
# Hybrid Coding Concept




Basis of every standard since H.261

CTB – Coding Tree Block
ME – Motion Estimation
PB – Prediction Block
Q – Quantization
TB – Transform Block
TR – Transform

# Performance history of standard generations



## Steps towards next generation standard – Versatile Video Coding (VVC)

- **Experimental software "Joint Exploration Model" (JEM) developed by JVET**
  - Intended to investigate potential for better compression beyond HEVC
  - Source code available from https://jvet.hhi.fraunhofer.de/
  - Was initially started extending HEVC software by additional compression tools, or replace existing tools (see next 3 pages)
- **Substantial benefit was shown over HEVC, both in subjective quality and objective metrics**
  - Proven in "Call for Evidence" (July 2017)
  - JEM was however not designed for becoming a standard (regarding all design tradeoffs)
  - Call for Proposals was issued by MPEG and VCEG (October 2017)
- **Call for Proposals very successful (responses received by April 2018)**
  - 46 category-specific submissions: 22 in SDR, 12 each in HDR and 360° video
  - All responses clearly better than HEVC, some evidently better than JEM
  - This marked the starting point for VVC development

# 2. Call for Proposals on Versatile Video Coding

**Recent developments in video compression standardization**

Jens-Rainer Ohm
Institute of Communication Engineering
RWTH Aachen University
ohm@ient.rwth-aachen.de

## Performance

- Submissions had to provide coded/decoded sequences
  - 4 rate points each, two constraint conditions "low delay" (LD) and "random access" (RA)
  - SDR: 5x HD (both LD and RA), 5x UHD-4K (only RA)
  - HDR: 5x HD (PQ grading), 3x UHD-4K (HLG grading)
  - 360°: 5 sequences 6K/8K for the full panorama

- Double stimulus test with two hidden anchors HEVC-HM & JEM
  - Rate points were defined such that lowest rate was typically less than "fair" quality for HEVC, but still possible to code
  - Quality was judged to be distinguishable when confidence intervals were non-overlapping

## Performance

- Measured by objective performance (PSNR), best performers report >40% bit rate reduction compared to HEVC, >10% compared to JEM (for SDR case)
  - Similar ranges for HDR and 360°
  - Obviously, proposals with more elements show better performance
  - Some proposals showed similar performance as JEM with significant complexity/run time reduction
  - 2 proposals used some degree of subjective optimization, not measurable by PSNR

- Results of subjective tests generally show similar (or even better) tendency
  - Benefit over HEVC very clear
  - Benefit over JEM visible at various points
  - Proposals with subjective optimization also showing benefit in some cases

## Performance compared to HEVC

- How often are best performing proposals *better* than HEVC at higher rate?
- Note: R1≅1 Mbit/s; R2 ≅1.6 Mbit/s; R3 ≅2.8 Mbit/s; R4 ≅4.6 Mbit/s

| $P_{best}$ vs HM | R1 vs R2 | R1 vs R3 | R1 vs R4 | R2 vs R3 | R2 vs R4 | R3 vs R4 |
|---|---|---|---|---|---|---|
| **SDR UHD** | 60% | 40% | 0% | 80% | 0% | 20% |
| **SDR HD/RA** | 40% | 0% | 0% | 20% | 0% | 20% |
| **SDR HD-/LD** | 40% | 0% | 0% | 0% | 0% | 0% |
| **HLG** | 67% | 0% | 0% | 67% | 0% | 33% |
| **PQ** | 40% | 0% | 0% | 40% | 0% | 20% |
| **360** | 40% | 20% | 0% | 20% | 0% | 60% |
| **Rate saving** | ~ 37.5% | ~ 65% | ~ 78% | ~ 43% | ~ 35% | ~ 39% |

## Performance compared to HEVC

- How often is HEVC *better* than best performing proposals at lower rate?
- Note: R1≅1 Mbit/s; R2 ≅1.6 Mbit/s; R3 ≅2.8 Mbit/s; R4 ≅4.6 Mbit/s

| HM vs $P_{best}$ | R1 vs R2 | R1 vs R3 | R1 vs R4 | R2 vs R3 | R2 vs R4 | R3 vs R4 |
|---|---|---|---|---|---|---|
| **SDR UHD** | 0% | 0% | 60% | 0% | 0% | 0% |
| **SDR HD/RA** | 0% | 60% | 100% | 0% | 80% | 0% |
| **SDR HD-/LD** | 0% | 60% | 80% | 0% | 80% | 0% |
| **HLG** | 0% | 0% | 100% | 0% | 67% | 0% |
| **PQ** | 0% | 60% | 100% | 0% | 60% | 0% |
| **360** | 0% | 40% | 80% | 0% | 40% | 0% |
| **Rate saving** | ~ 37.5% | ~ 65% | ~ 78% | ~ 43% | ~ 65% | ~ 39% |

# 3. Tools for improved compression

**Recent developments in video compression standardization**

Jens-Rainer Ohm
Institute of Communication Engineering
RWTH Aachen University
ohm@ient.rwth-aachen.de

## What was proposed?

- In terms of large architecture: Most proposals similar, no deviation from hybrid coding mainstream
- Most improvements from further refinements of well-known building blocks of HEVC and JEM
  - Partititioning: Multi-type tree (Quad/binary/ternary), and finer
  - Intra prediction using
    - directional modes, DC and planar
    - sample smoothing with various adaptation methods
    - inheritance of chroma modes and chroma sample prediction from luma
  - Inter prediction using advanced motion vector prediction, affine models, sub-block partitioning
  - Switchable primary transforms, mostly DCT/DST variants
  - Secondary transforms targeting specific prediction residual characteristics
  - Adaptive loop filter based on local classification, some new variants
  - Quantization / context-adaptive arithmetic coding

## What was proposed?

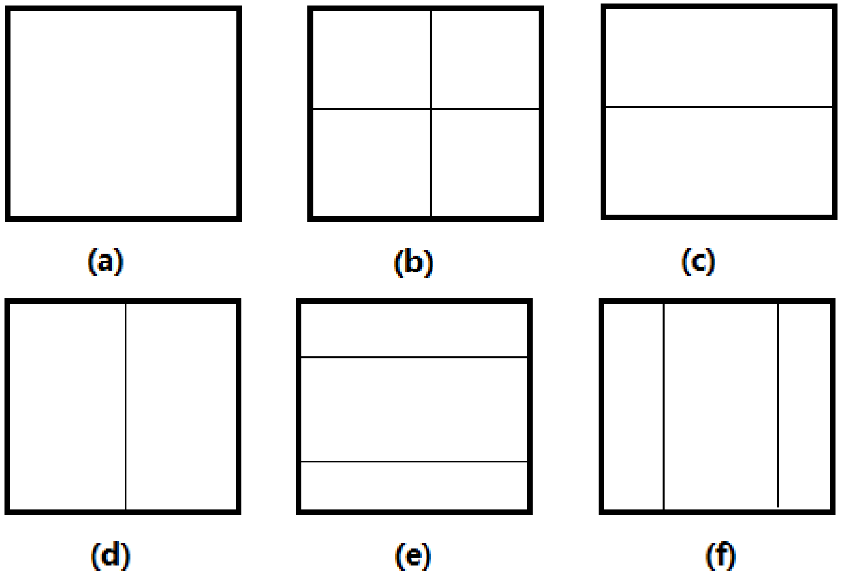
- Compression-improving tools:
  - Template matching tools (decoder side) for purposes of mode/MV derivation and sample prediction both in intra and inter coding
  - Finer partitioning: Asymmetric rectangular, geometric/wedge
  - Enlarged intra reference area & intra block copy
  - Additional non-linear, de-noising and statistics-based loop filters
  - **Neural networks for prediction, loop filtering, upsampling**

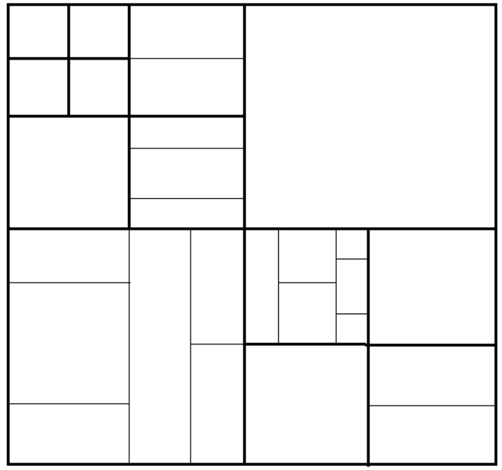
- HDR specific:
  - New adaptive reshaping and quantization, also in-loop
  - HDR-specific modifications of existing tools, e.g. deblocking

- 360-video specific:
  - Variants of projection formats, geometry-corrected face boundary padding
  - Modification and disabling of existing tools at face boundaries

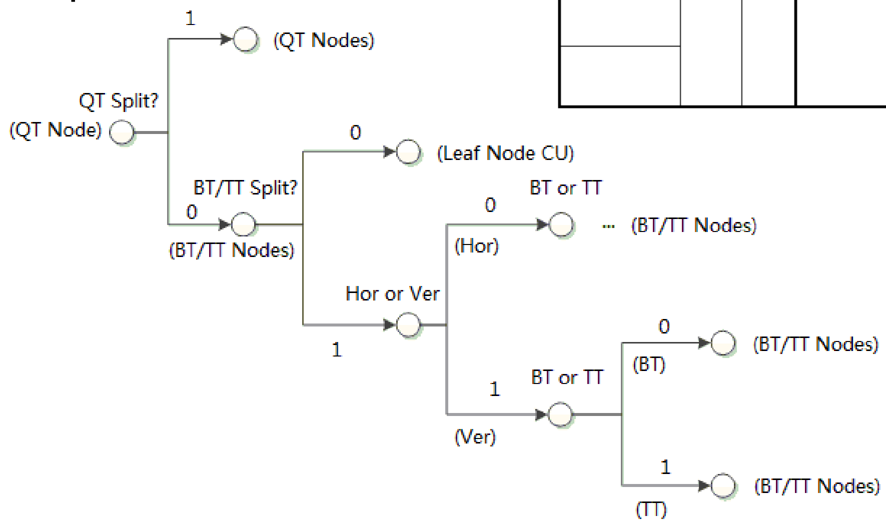
# New trend: More flexible block splitting

- Simple multi-type tree split was used in several proposals, can be alternated as ternary/binary split (originating from quadtree leaf)



Example:




(source: JVET-J1002)

- Further proposed variants of partitioning included
  – Asymmetric rectangular binary split modes
  – Diagonal (wedge-shaped) binary split modes

## VVC Test Model and Benchmark Set

- **VVC Working Draft 1 / Test Model 1 (VTM1):** basic approach built on "reduced HEVC" starting point
- **VTM Block structure**
  - Unified multi-type tree (binary/ternary splits after quad-tree, coding block unites prediction and transform)
  - CTU size 128x128, rectangular blocks (dyadic sizes), smallest luma size 4x4
  - Maximum transform size 64x64
- **VTM: Some removed elements of HEVC:**
  - Mode dependent transform (DST-VII), mode dependent scan
  - Strong intra smoothing
  - Sign data hiding in transform coding
  - Unnecessary high-level syntax (e.g. VPS)
  - Tiles and wavefront
  - Quantization weighting

- **Benchmark Set** defined in addition to VTM, including the following well-known JEM tools:
  - 65 intra prediction modes
  - Coefficient coding
  - AMT + 4x4 NSST
  - Affine motion
  - Geometry transformation based adaptive loop filter (GALF)
  - Subblock merge candidate (ATMVP)
  - Adaptive motion vector precision
  - Decoder motion vector refinement
  - LM Chroma mode

  Purpose: testing benefit of technology against better performing set

# 4. Methods related to deep learning

**Recent developments in video compression standardization**

Jens-Rainer Ohm
Institute of Communication Engineering
RWTH Aachen University
ohm@ient.rwth-aachen.de

# NN for Video

- NN-based approaches were so far more successful in still image coding rather than video coding
  - Perceptual criteria also better understood for images
- In video coding, motion compensation is a most effective key component
  - Requires motion estimation for which "conventional" algorithms appear to be less complex
  - Analogy: Eye tracking – the brain processes a motion compensated input
- CNN have been demonstrated to provide benefit in context of video coding for
  - Resolution up-conversion
  - Post-processing and loop filtering
  - Intra coding
  - Encoder optimization, in particular partitioning which is basically a segmentation problem

# CNN for Resolution up-conversion

- Basic idea of dynamic resolution coding:
  - Downsample and code by lower resolution (less bitrate cost)
  - Upsample at decoder side to full resolution
  - Encoder decides using full res, conventional or CNN-based down- and upsampling
  - CNN-based could generate super-resolution upsampling, sharper edges, etc.
- Can be implemented in combination with intra and inter prediction coding
- Operated on block by block basis

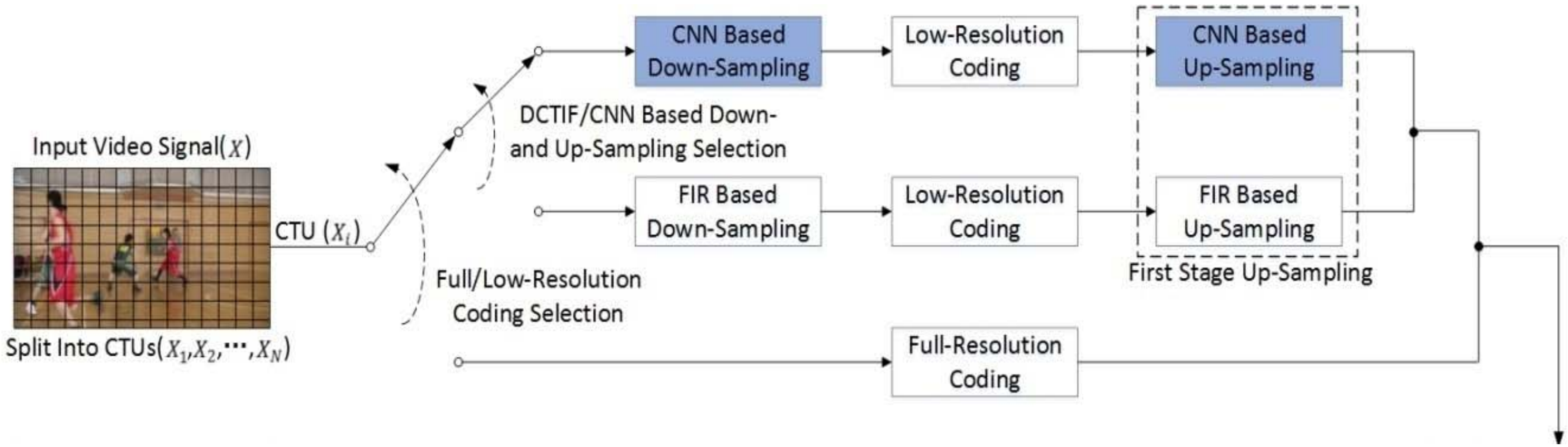
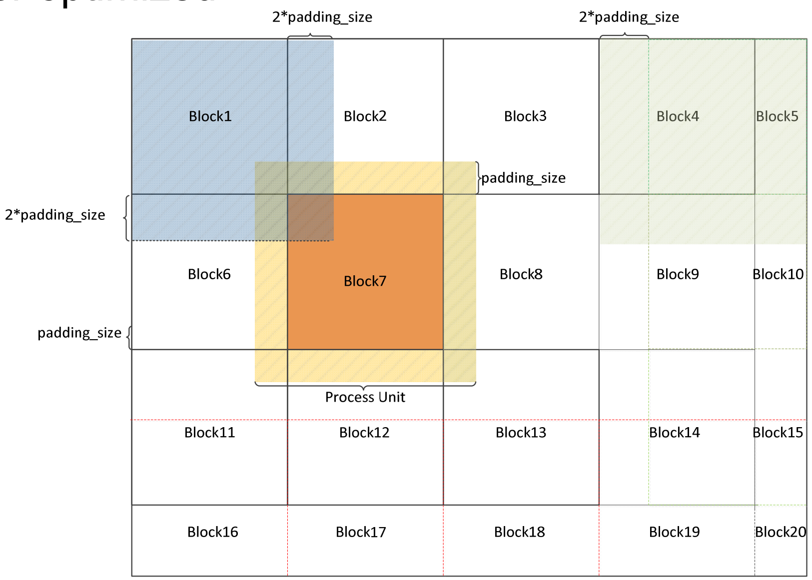
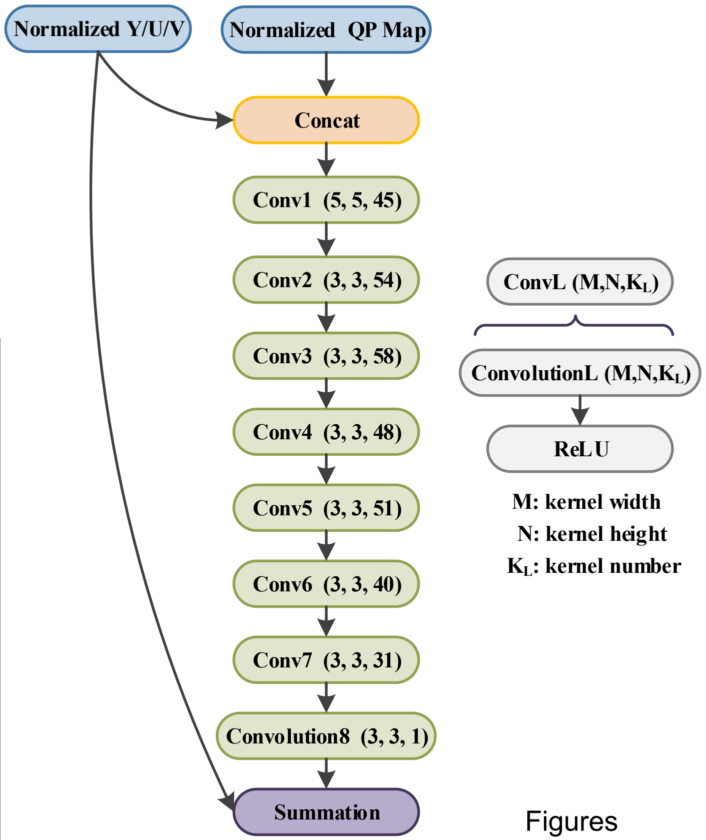



Figure from JVET-J0032

## CNN for Loop filtering

- Loop filtering is common in video coding
  - removes compression artifacts from reconstruction
  - improves prediction  from reconstructed frames
- Generally, signal-adaptive and non-linear filters
  - e.g., de-blocking, de-ringing, de-banding
  - edge-adaptive & Wiener optimized
  - bi-lateral filters
  - ...
- CNN reconstruction provides additional gain (3-5% rate red.) and might replace some conventional filters
- Can be operated on block basis, parallel processing possible




Figures from JVET-I0022

## Neural networks for intra prediction

- Neural networks were demonstrated to provide improved intra prediction, compared to conventional directional and planar modes
- Mostly fully connected networks have been used for this purpose (no convolutional layers)
- Average rate reductions of 4-5% (for intra coding) have been reported
- Examples of prediction demonstrate the benefit of non-linear processing

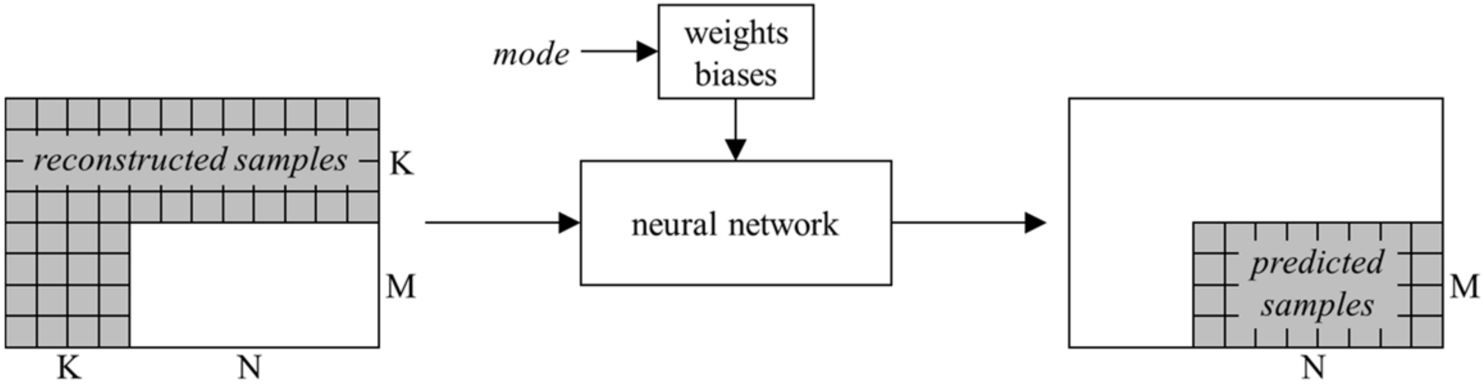



Figure
from JVET-J0037

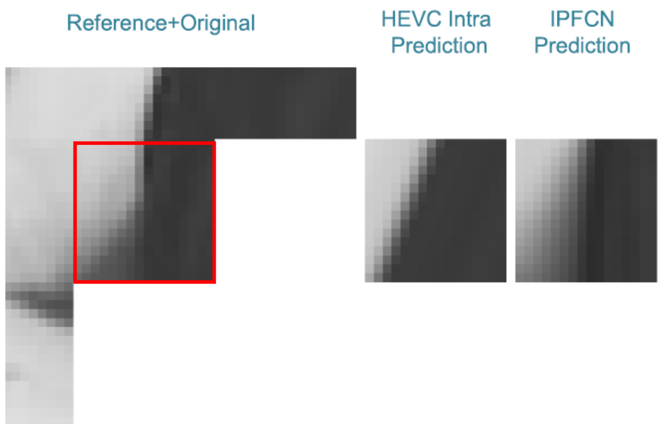



Figure
from Li et al.
IEEE-TCSVT
July 2018

## Summary and Outlook

- Video compression is a lively area of research, major and ongoing progress in standardization
- AVC became the most widely used standard worldwide
- HEVC has demonstrated significant technical and performance advance and is currently ramping up in markets
- The work of JVET has demonstrated that significant improvement of compression beyond HEVC is possible
  - Development of experimental JEM platform demonstrated initial benefit
  - Successful Call for Proposals unveiled that even better performance is possible
  - First steps towards VVC by establishing a first draft text and test model
- This is only the beginning
  - Additional benefit may come from other emerging technology, e.g. deep learning
  - Goal of 50% bit rate reduction with same quality as HEVC can probably be reached
  - Rigid process necessary to establish a reasonable tool combination

## Further Information

- Document archives (publicly accessible)
  - http://phenix.it-sudparis.eu/jct
  - http://phenix.it-sudparis.eu/jvet
  - http://ftp3.itu.ch/av-arch/jctvc-site
  - http://ftp3.itu.ch/av-arch/jvet-site

- Software for VTM, HEVC, JEM, and 360 Video (publicly accessible):
  - https://jvet.hhi.fraunhofer.de/svn/svn_VVCSoftware_<VTM|BMS>
  - https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/
  - https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/
  - https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/