Versatile Video Coding –
towards the next generation of video compression

Picture Coding Symposium 2018, San Francisco, 2018-06-26
"Bridging the Gap" Invited Talk

Jens-Rainer Ohm, Institute of Communication Engineering, RWTH Aachen University
Gary J. Sullivan, Microsoft Corporation

Outline

1. Introduction and history of video coding standardization
2. Call for Proposals on Versatile Video Coding – results
3. Tools for improved compression – some details
4. Next steps, summary and outlook

1. Introduction and history of video coding standardization

Versatile Video Coding – towards the next generation of video compression

PCS 2018 – "Bridging the Gap" Invited Talk

Jens-Rainer Ohm and Gary Sullivan

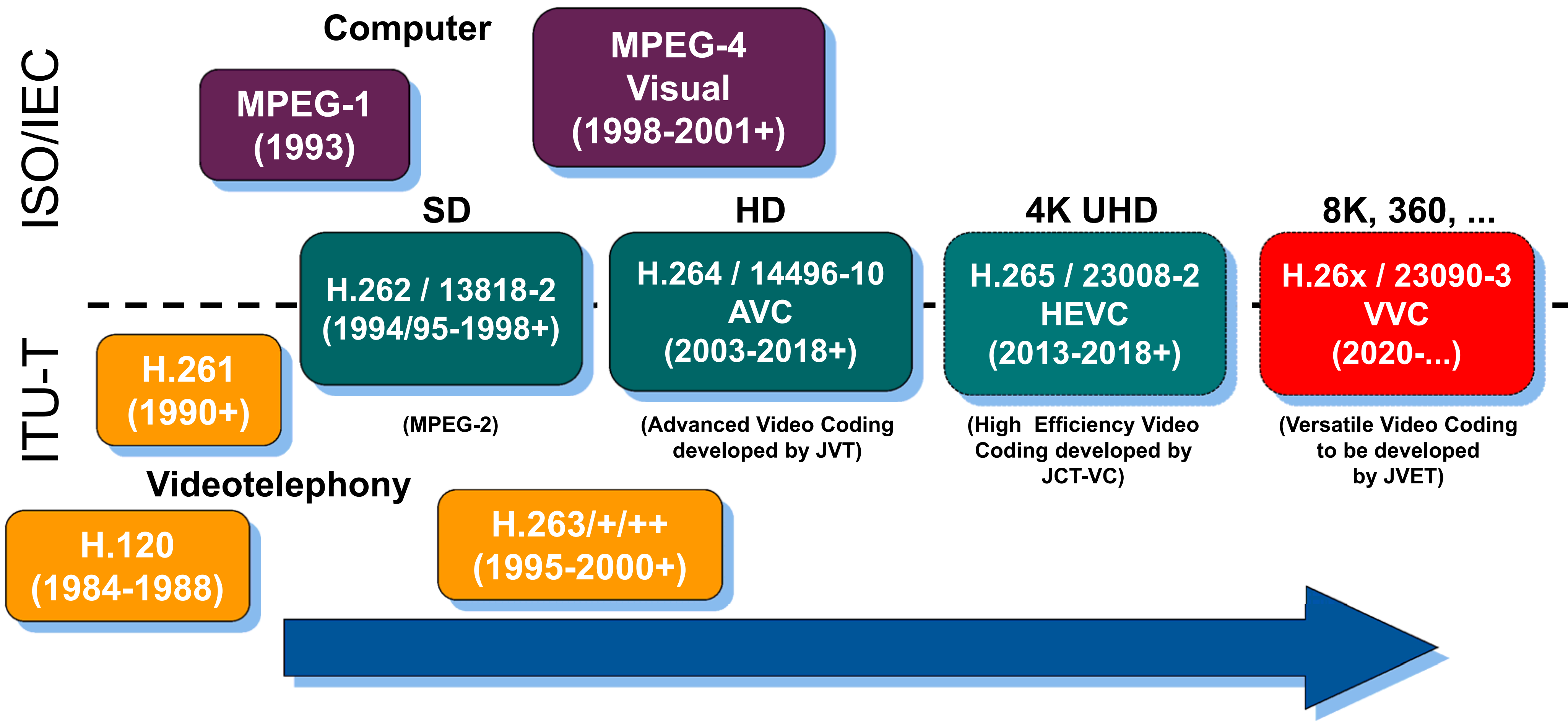
Motivation for permanent improvements in video compression

- **Video is already the vast majority of data traffic (~75%)**
- **Video is continually increasing by resolution**
  - HD existing, UHD (4Kx2K, 8Kx4K) appearing
  - Mobile services going towards HD/UHD
  - Stereo, multi-view, 360° video
- **Devices available to record, display and distribute ultra-high resolutions**
  - Becoming affordable for home and mobile consumers
  - Surveillance uses expanding, with remote storage and analysis
- **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 grows faster than feasible network transport capacities**
  - Better video compression (e.g. 50% rate of current HEVC) needed, even after availability of 5G

Video coding standardization organisations

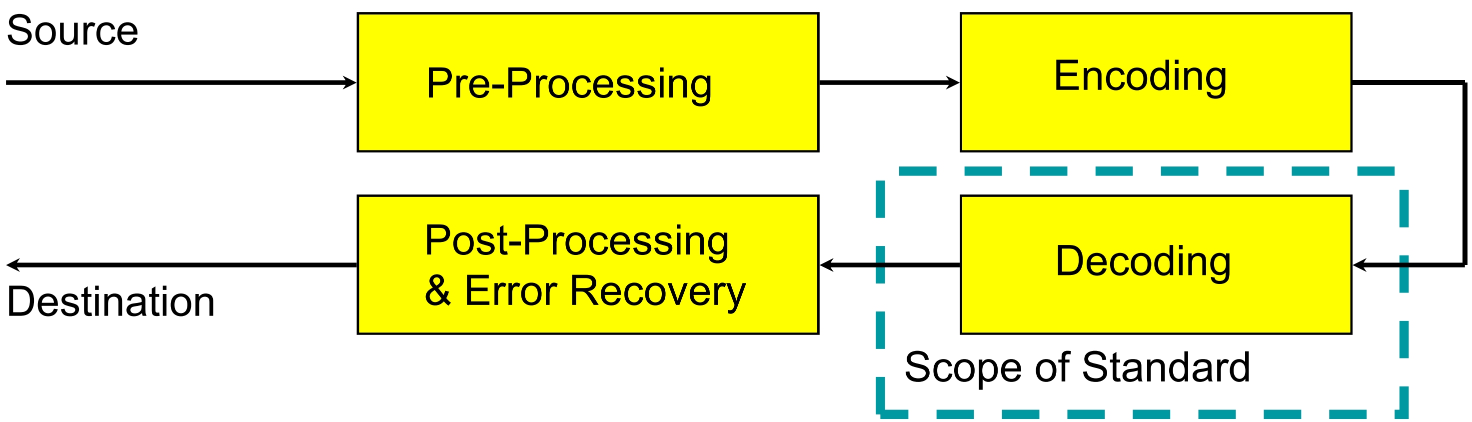
- **ISO/IEC MPEG** = "**Moving Picture Experts Group**"
  (ISO/IEC JTC 1/SC 29/WG 11 = International Standardization Organization and International Electrotechnical Commission, Joint Technical Committee 1, Subcommittee 29, Working Group 11)
- **ITU-T VCEG** = "**Video Coding Experts Group**"
  (ITU-T 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)

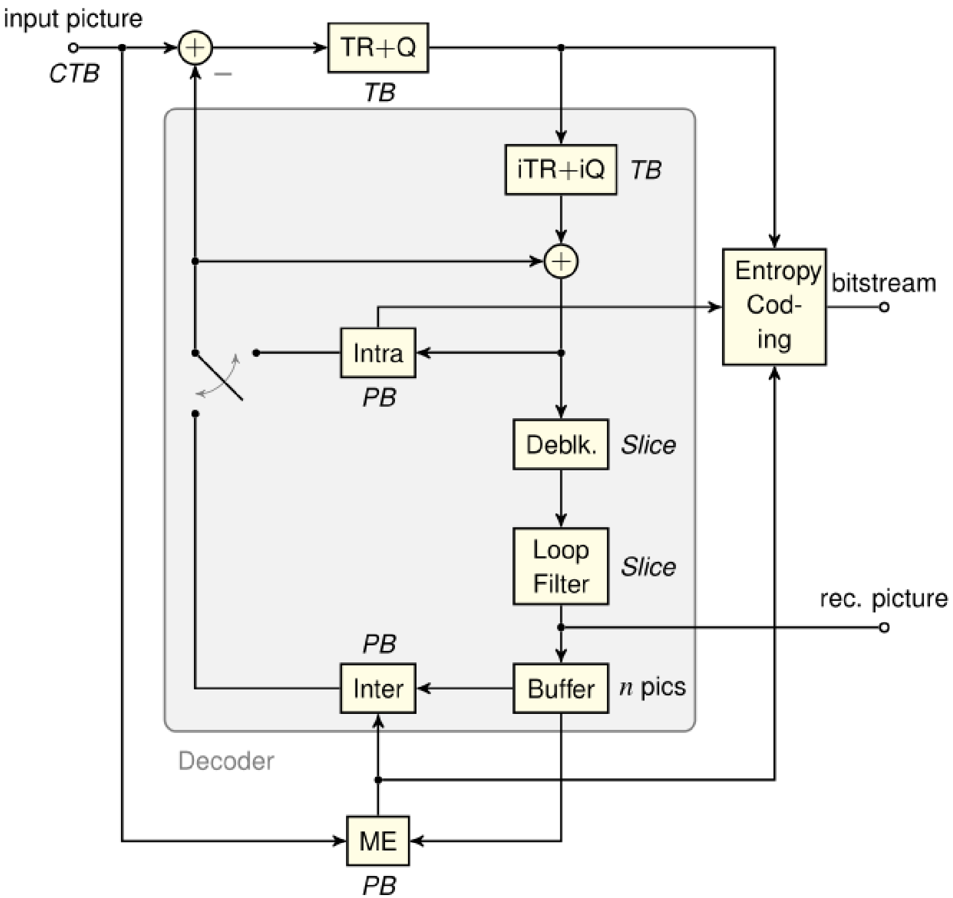


The scope of video standardization

- **Only Specifications of the *Bitstream*, *Syntax,* and *Decoder* are standardized:**
  - Permits optimization beyond the obvious
  - Permits complexity reduction for implementability
  - Provides *no* guarantees of quality



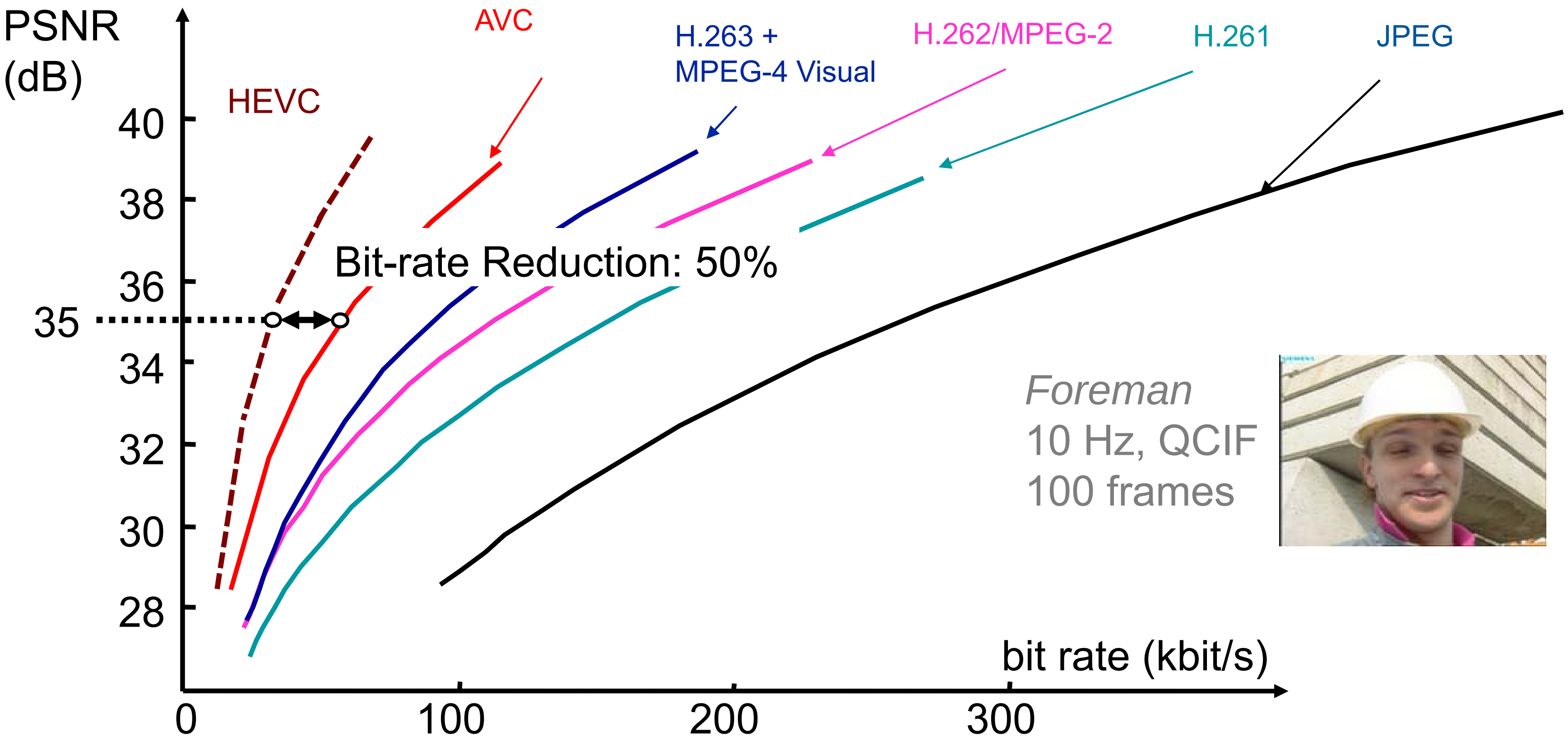
Hybrid Coding Concept



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



What made this happen over the years?

- **Improvements of motion compensation**
  - Variable partitions & merged partitions
  - Flexible frame referencing & combined prediction
  - Sub-sample precision and high performance sub-sample interpolation
  - More efficient vector prediction & coding, supporting large vector ranges
- **Improvements of 2D coding**
  - Efficient intra prediction and intra mode coding
  - Design of transform bases and variable transform block sizes
- **Loop filtering for artifact reduction**
  - Deblocking, sample-adaptive offset
- **Improvements of entropy coding**
  - Flexible binarization of syntax elements
  - Arithmetic coding
  - Adaptation and usage of context information
- **These are coupled with encoder optimization**
  - Rate distortion optimization – spend bits where they give best benefit in terms of distortion reduction
  - Adaptive rate control and perceptually tuned quantization

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)**
  - 32 companies in 21 proponent groups responded
  - 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

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

- **What does "Versatile" stand for?**

- **VVC should be usable for many types of data**
  - SDR and HDR up to extreme high resolutions
  - All kind of camera generated content
  - Computer generated content
  - Non-camera video modalities e.g. medical data
  - 360°, lightfield, depth, and volumetric video

- **VVC should support flexible random and localized access**
- **VVC should be easily configurable for various application needs**
- **The core of VVC should consist only of minimum amount of necessary (and well-understood) building blocks – "clean slate" design as compared to HEVC**

2. Call for Proposals on Versatile Video Coding – results

Versatile Video Coding – towards the next generation of video compression

PCS 2018 – "Bridging the Gap" Invited Talk

Jens-Rainer Ohm and Gary Sullivan

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

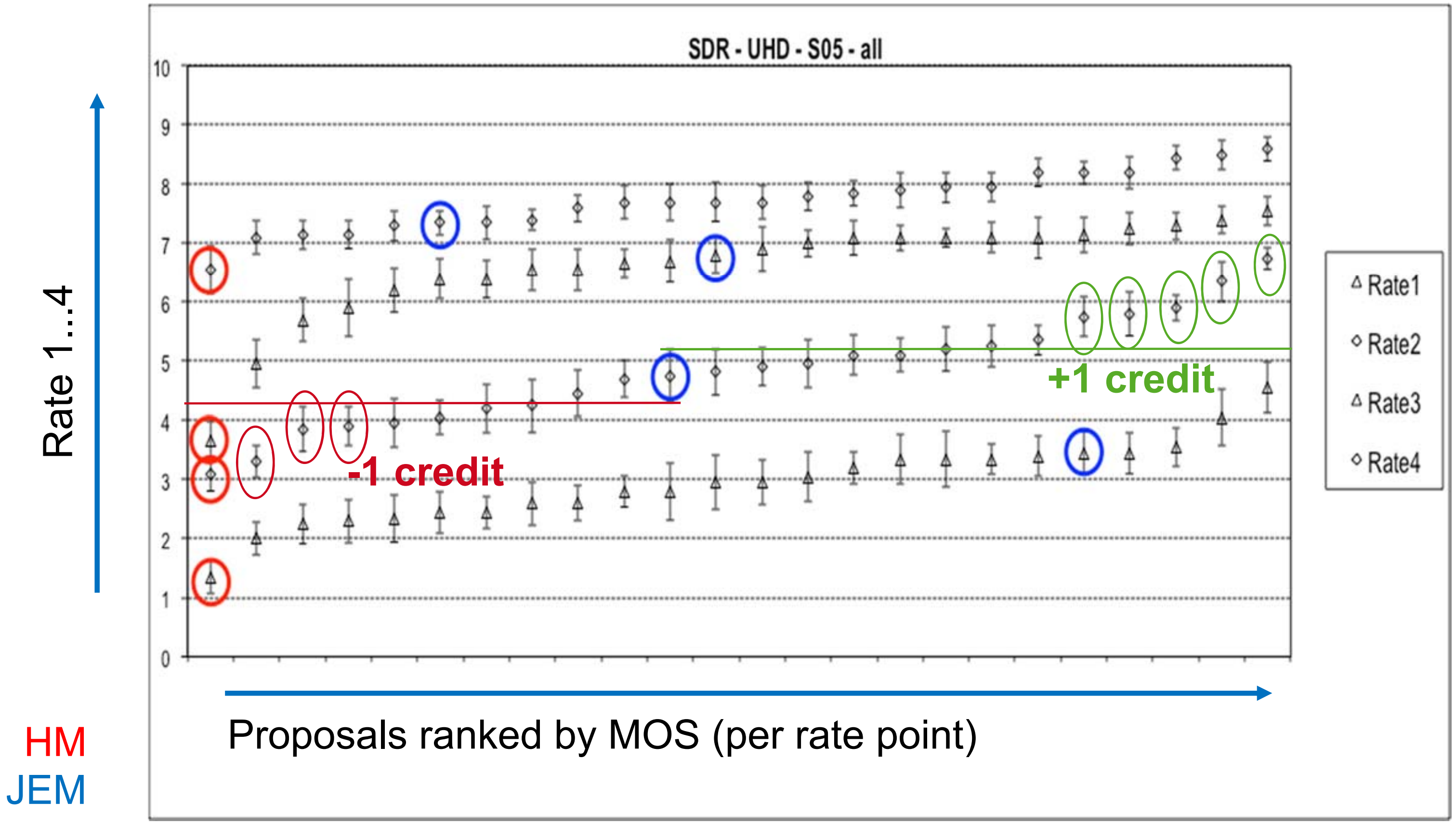
Performance

- 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

- JVET-J1003:
  Report of subjective evaluation contains 28 plots as shown, one per sequence
- Count significant cases of positive/ negative benefit with non-overlapping confidence interval against JEM



Performance SDR

- "Mean" and "significance-count" method suggested at least 7 proposals that were obviously better than JEM at same rate
  - Approx. 0.5 better in MOS
  - Proven subjectively better in particular for lower rates

Mean MOS

Pxx	6,53	Pnn	6,04
Pxx	6,46	Pnn	6,04
Pxx	6,41	Pnn	6,03
Pxx	6,37	Pnn	6,03
Pxx	6,33	Pnn	6,01
Pxx	6,33	JEM	6,01
Pxx	6,26	Pnn	6,00
Pnn	6,23	Pnn	5,96
Pnn	6,17	Pnn	5,94
Pnn	6,15	Pnn	5,88
Pnn	6,13	Pnn	5,86
Pnn	6,11	HM	4,57

–60 ... +60

Significance vs. JEM

Pxx	10	Pnn	1
Pxx	8	JEM	0
Pxx	8	Pnn	0
Pxx	6	Pnn	-1
Pxx	6	Pnn	-1
Pxx	6	Pnn	-1
Pxx	6	Pnn	-2
Pnn	3	Pnn	-2
Pnn	3	Pnn	-2
Pnn	2	Pnn	-3
Pnn	2	Pnn	-4
Pnn	1	HM	-36

Performance HDR / 360°

- Similar tendency in HDR and 360° categories
- Mostly same coding tools as in SDR provide good benefit

HDR — Mean MOS / −32 ... +32 Signif. vs. JEM

Mean MOS		Signif. vs. JEM	
Pxx	6,04	Pxx	7
Pxx	6,00	Pxx	3
Pxx	5,94	Pxx	2
Pxx	5,93	Pxx	2
Pxx	5,86	Pxx	2
Pnn	5,85	Pnn	1
Pnn	5,80	Pnn	1
Pnn	5,67	JEM	0
JEM	5,62	Pnn	0
Pnn	5,60	Pnn	0
Pnn	5,59	Pnn	-1
Pnn	5,45	Pnn	-1
Pnn	5,11	Pnn	-6
HM	4,14	HM	-20

360° — Mean MOS / −20 ... +20 Signif. vs. JEM

Mean MOS		Signif. vs. JEM	
Pxx	6,20	Pxx	9
Pxx	6,19	Pxx	9
Pxx	6,06	Pxx	8
Pxx	6,03	Pxx	7
Pxx	5,99	Pxx	7
Pxx	5,96	Pnn	6
Pxx	5,86	Pxx	5
Pnn	5,69	Pxx	4
Pnn	5,67	Pnn	2
Pnn	5,51	Pnn	1
Pnn	5,45	Pnn	1
JEM	5,11	JEM	0
HM	3,79	HM	-9
Pnn	3,45	Pnn	-12

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: 1-xx% means that best performing proposal is equal or better
- 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%

Performance compared to HEVC

- The subjective quality of best performing proposals is *always equal or sometimes better* (~1/3 of cases) than HEVC at next higher rate point, over all categories (with approx. 40% less rate)
- The subjective quality of best performing proposals is always equal or sometimes better (~1/5 of cases) than HEVC at 2nd higher rate point, in SDR-UHD category (with approx. 65% less rate)
- Though it is not always the same proposal that performs best at a given rate point, it can be anticipated that merits of different proposals can be combined
- 50% (or more) bit rate reduction with same quality will probably be achievable by the new standard

3. Tools for improved compression – some details

Versatile Video Coding – towards the next generation of video compression

PCS 2018 – "Bridging the Gap" Invited Talk

Jens-Rainer Ohm and Gary Sullivan

What is JEM?

- Exploration software that was built on top of HEVC-HM, with modifications / added elements as below
  - Not optimized for performance/complexity tradeoff (but need for keeping manageable encoder runtime)

- Block structure
  - Larger Coding Tree Unit (up to 256x256) and transforms (up to 64x64)
  - Quadtree plus binary tree (QTBT) block structure replaced quadtree structure of HEVC

- Intra prediction tools
  - 65 intra prediction directions
  - 4-tap interpolation filter for intra prediction
  - Boundary filter applied to other directions in addition to horizontal and vertical ones
  - Cross-component linear model (CCLM) prediction
  - Position dependent intra prediction combination (PDPC)

- Transform
  - Explicit multiple core transform
  - Mode dependent non-separable secondary transforms in intra coding

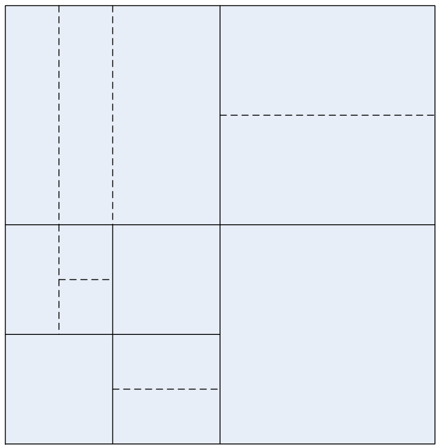
What is JEM? (cont.)

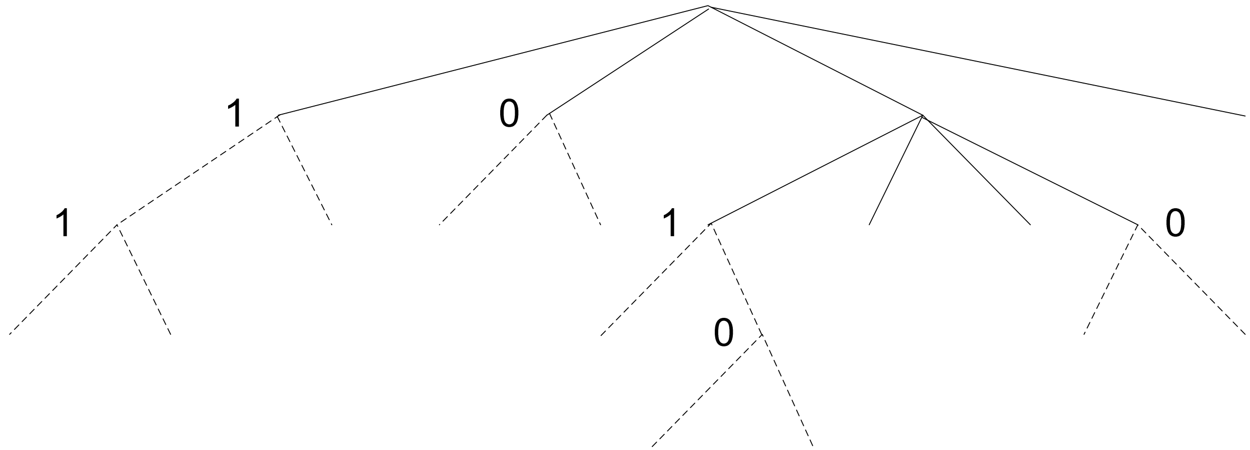
- Inter prediction
  - Subblock level motion vector prediction
  - Locally adaptive motion vector resolution (AMVR)
  - 1/16 pel motion vector storage accuracy
  - Overlapped block motion compensation (OBMC)
  - Local illumination compensation (LIC)
  - Affine motion prediction
  - Pattern matched motion vector derivation at decoder
  - Bi-directional optical flow (BIO) for improved motion compensation at decoder

- In-loop filters
  - Adaptive loop filter (ALF)
  - Bilateral filter

- CABAC design
  - Context model selection for transform coefficient levels
  - Multi-hypothesis probability estimation, improved initialization for context models

JEM: Quad-Tree plus Binary Tree Partitioning (QTBT)



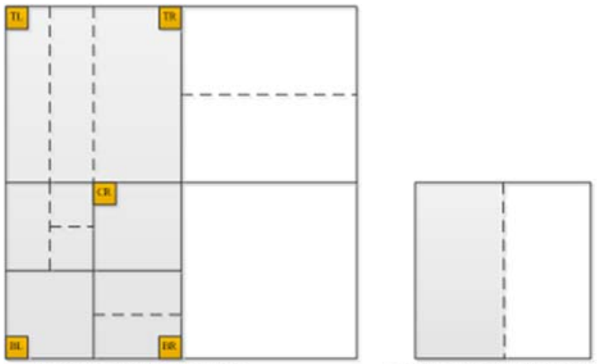


- QTBT uses same rectangular block partitioning into coding units (CU) for mode signalling, prediction and transform
- QTBT structure starts from 128 × 128 square blocks (optionally larger or smaller) and always performs square-block split
- Binary trees starting from leaves of quad-tree (with horizontal / vertical split indication)
  → final CU can have either square or rectangular shape

Figure from: Jianle Chen et al. *Algorithm Description of Joint Exploration Test Model 7*. Doc. JVET-G1001. Torino, IT, 7th meeting: Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Jul. 2017.

JEM Intra Prediction Modes

- Concept of HEVC as basis
  - Higher number of prediction modes
  - Larger maximum block size
  - Additional position and mode dependent filtering/smoothing
- Chroma
  - Prediction modes from neighbors
  - Derived modes from collocated luma



(a) Luma QTBT structure (b) Chroma QTBT structure

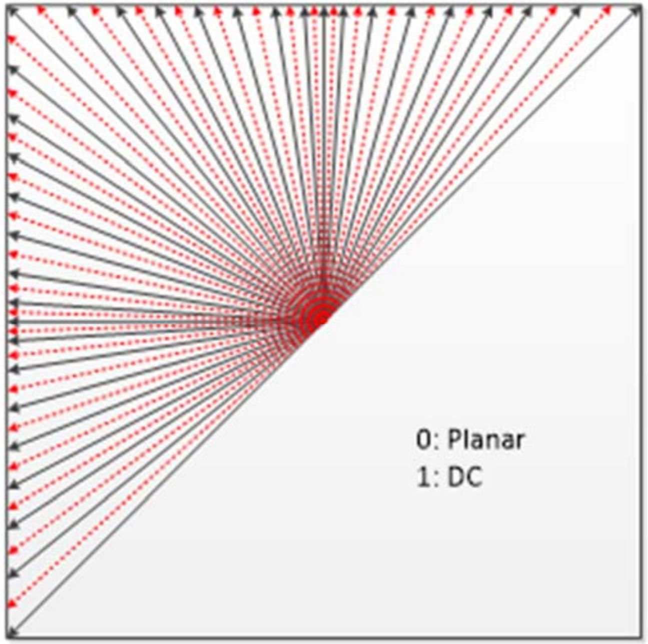
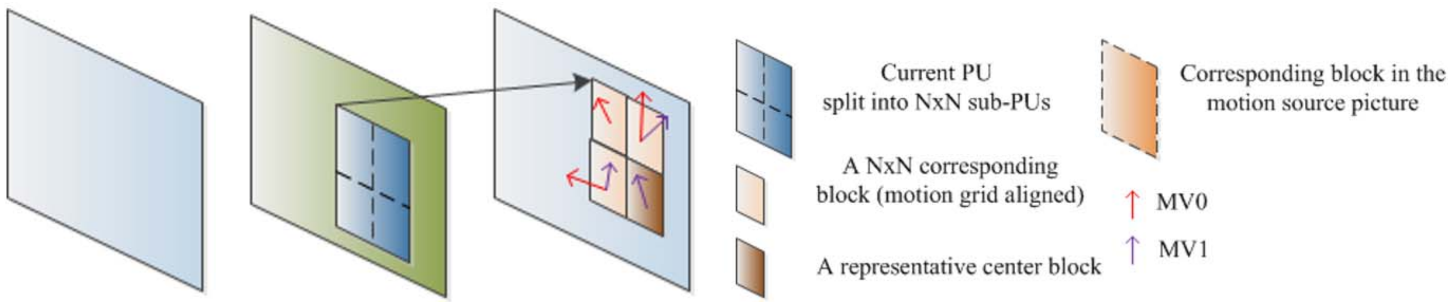


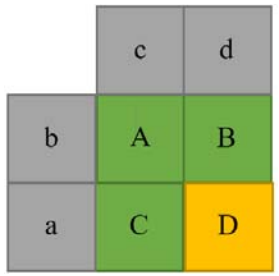
Figure from: Jianle Chen et al. Algorithm Description of Joint Exploration Test Model 7. Doc. JVET-G1001. Torino, IT, 7th meeting: Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Jul. 2017.

JEM: Sub-CU based motion vector prediction

- CU: at most one set of motion parameters for each prediction direction
- Option to split large CU into sub-CUs
  - Alternative temporal motion vector prediction (ATMVP)
    - Fetch multiple sets of motion information from multiple blocks in collocated reference picture



  - Spatial-temporal motion vector prediction (STMVP)
    - Derive recursively by temporal motion vector predictor and spatial neighbouring motion vector



- ATMVP and STMVP: additional merge candidates (list extended to max 7)

Figures from: . JVET-G1001: *Algorithm Description of Joint Exploration Test Model 7*.

JEM: Affine Motion Vector Derivation for MC

- Motion vector field (MVF) for CU, applicable MV derived for each 4 × 4 block at 1/16 pel resolution
  - Control point motion vector (CPMV)

$$\begin{cases} v_x = \dfrac{(v_{1x} - v_{0x})}{w} x - \dfrac{(v_{1y} - v_{0y})}{w} y + v_{0x} \\ v_y = \dfrac{(v_{1y} - v_{0y})}{w} x + \dfrac{(v_{1x} - v_{0x})}{w} y + v_{0y} \end{cases}$$

- AF INTER mode
  - Signalling CPMV difference from predictor
  - Block width and height ≥ 8 required
- AF MERGE mode
  - Derivation of CPMV from neigborhood

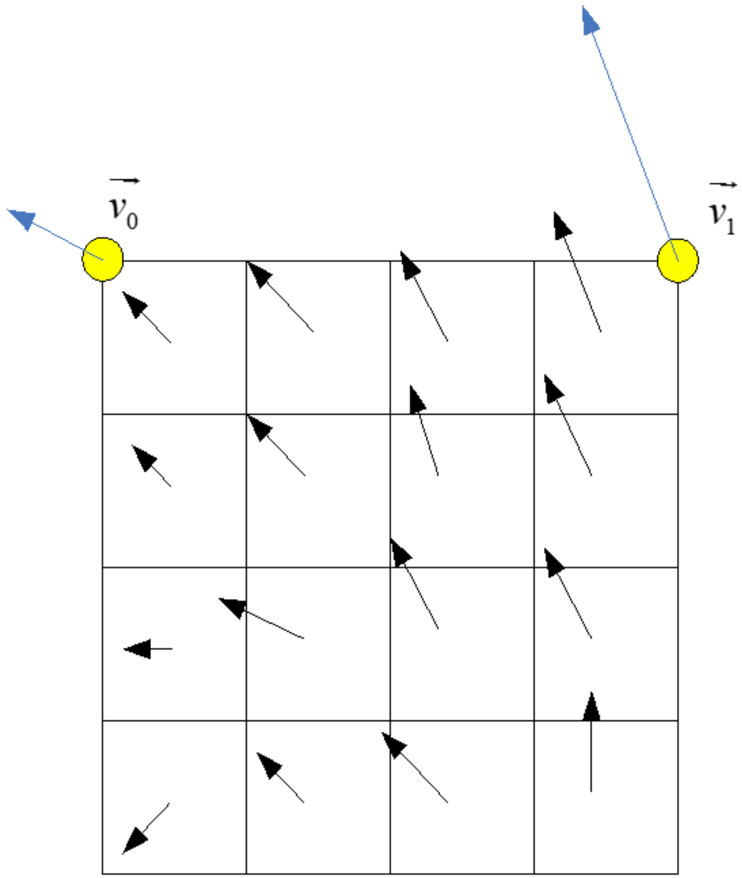
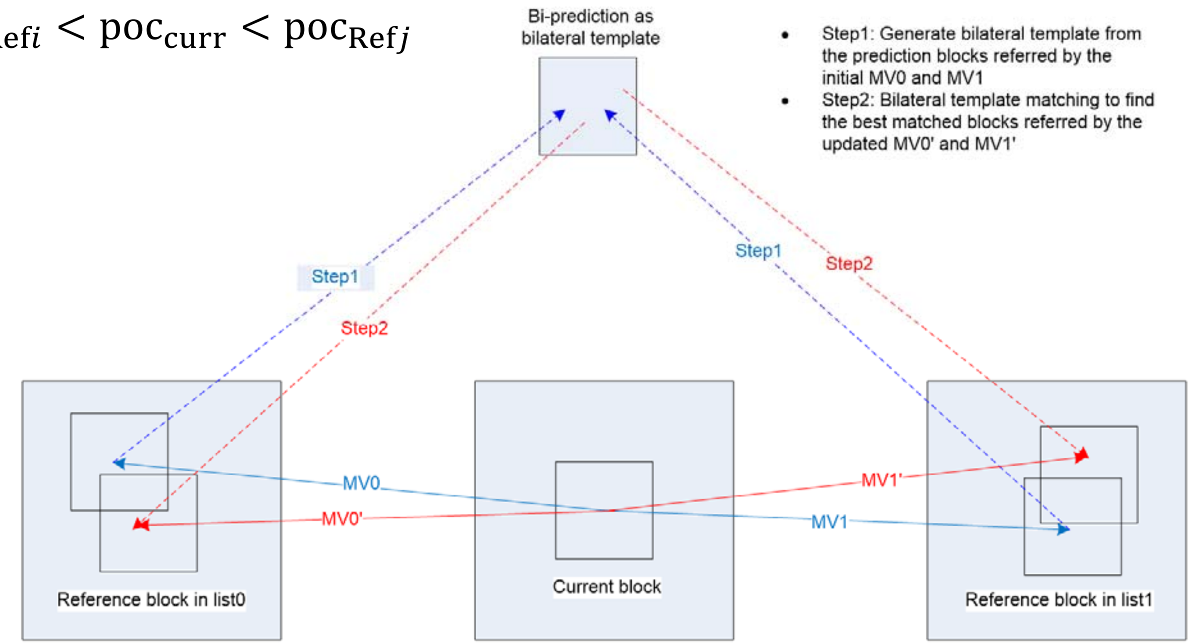


Figure from: JVET-G1001: *Algorithm Description of Joint Exploration Test Model 7*.

JEM: Decoder-side Motion Vector Refinement (DMVR)

- MVs of bi-prediction refined by bilateral template matching process
- Search between bilateral template and reference pictures
  ⇒ refined MV without further signaling
- Applied only with reference pictures with $\mathrm{poc}_{\mathrm{Ref}i} < \mathrm{poc}_{\mathrm{curr}} < \mathrm{poc}_{\mathrm{Ref}j}$
- Not applied if enabled in CU:
  - LIC,
  - Affine motion,
  - FRUC, or
  - sub-CU merge candidate



Figures from: JVET-G1001: *Algorithm Description of Joint Exploration Test Model 7*.

JEM: Cross-Component Linear Model Prediction (CCLM)

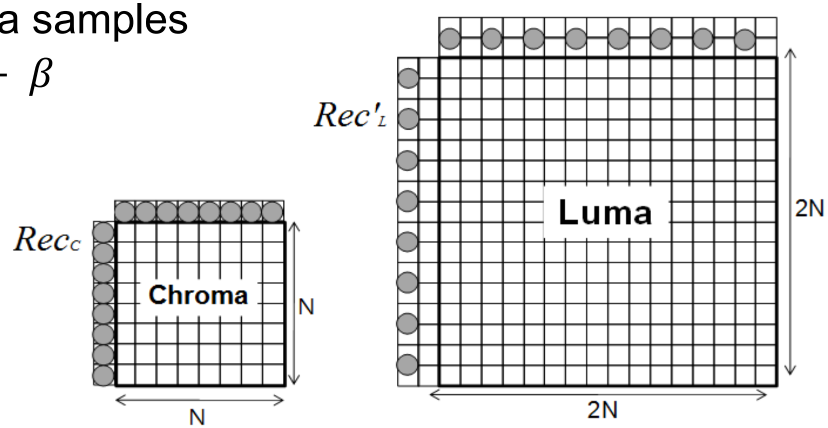
- Chroma samples predicted using corresponding reconstructed luma samples

$$pred_C(i,j) = \alpha \cdot rec_L{}'(i,j) + \beta$$

- Parameters α and β: minimize regression error between neighbouring reconstructed luma and chroma samples around current block

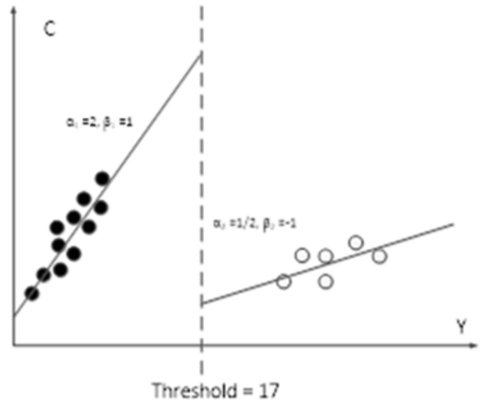
- Further prediction between chroma components with updated parameters

$$pred^*_{Cr}(i,j) = pred_{Cr}(i,j) + \alpha \cdot resi_{Cb}{}'(i,j)$$



Multiple model CCLM mode (MMLM)

- Neighbouring luma samples and neighbouring chroma samples classified into two groups
- Linear model for each group



Figures from: JVET-G1001: *Algorithm Description of Joint Exploration Test Model 7*.

JEM Transforms

Tables from: JVET-G1001: *Algorithm Description of Joint Exploration Test Model 7*.

- **Large block-size transforms** with high-frequency zeroing
  - Maximum transform size up to 128×128
  - Coefficients with column / row index > 32 set to 0 if
    - Block width > 64
    - Block height > 64, respectively

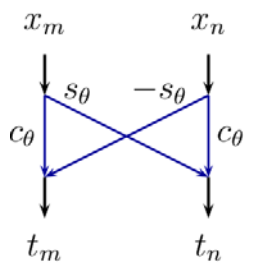
- **Adaptive multiple core transform** (AMT)
  - Transform matrices quantized more accurately
  - Applicable for block sizes $\leq 64 \times 64$
  - Indicated by CU flag
  - Mode-dependent transform-set selection for intra prediction modes

Transform Set	Transform Candidates
0	DST-VII, DCT-VIII
1	DST-VII, DST-I
2	DST-VII, DCT-VIII

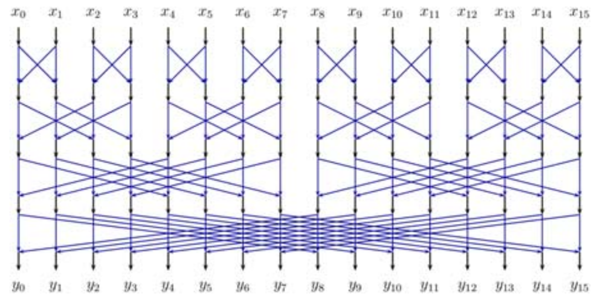
Transform Type	Basis function $T_i(j)$, $i, j=0, 1,\ldots, N-1$
DCT-II	$T_i(j) = \omega_0 \cdot \sqrt{\frac{2}{N}} \cdot \cos\left(\frac{\pi \cdot i \cdot (2j+1)}{2N}\right)$ where $\omega_0 = \begin{cases} \sqrt{\frac{2}{N}} & i = 0 \\ 1 & i \neq 0 \end{cases}$
DCT-V	$T_i(j) = \omega_0 \cdot \omega_1 \cdot \sqrt{\frac{2}{2N-1}} \cdot \cos\left(\frac{2\pi \cdot i \cdot j}{2N-1}\right)$, where $\omega_0 = \begin{cases} \sqrt{\frac{2}{N}} & i = 0 \\ 1 & i \neq 0 \end{cases}$, $\omega_1 = \begin{cases} \sqrt{\frac{2}{N}} & j = 0 \\ 1 & j \neq 0 \end{cases}$
DCT-VIII	$T_i(j) = \sqrt{\frac{4}{2N+1}} \cdot \cos\left(\frac{\pi \cdot (2i+1) \cdot (2j+1)}{4N+2}\right)$
DST-I	$T_i(j) = \sqrt{\frac{2}{N+1}} \cdot \sin\left(\frac{\pi \cdot (i+1) \cdot (j+1)}{N+1}\right)$
DST-VII	$T_i(j) = \sqrt{\frac{4}{2N+1}} \cdot \sin\left(\frac{\pi \cdot (2i+1) \cdot (j+1)}{2N+1}\right)$

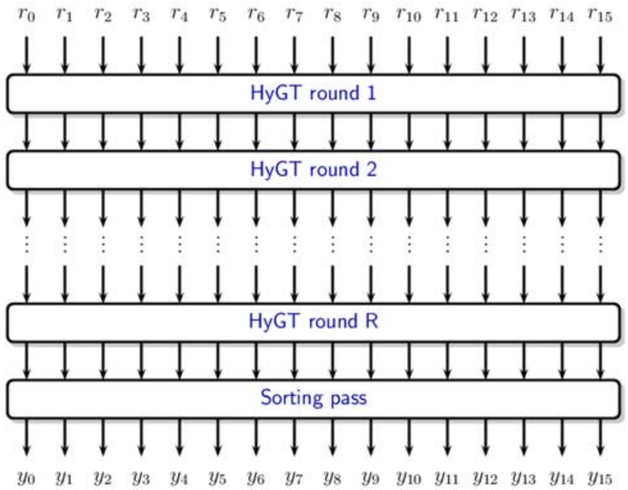
JEM: Mode-Dependent Non-separable Secondary Transforms (MDNSST)

- Motivation
  - Remaining correlation between coefficients after primary transform!
  - Dependency on intra prediction mode!
- Only applied to the low frequency coefficients after the primary transform
  - For blocks $\geq 8 \times 8$, application of 8×8 tranform to lowest frequency coefficients of primary transform
  - For blocks $< 8 \times 8$, application of 4×4 tranform to lowest frequency coefficients of primary transform
- Implementation by Hypercube-Givens Transform (HyGT)
- Two rounds for 4×4, four rounds for 8×8 secondary transforms



$$t_m = x_m \cos\theta - x_n \sin\theta$$
$$t_n = x_m \sin\theta + x_n \cos\theta$$

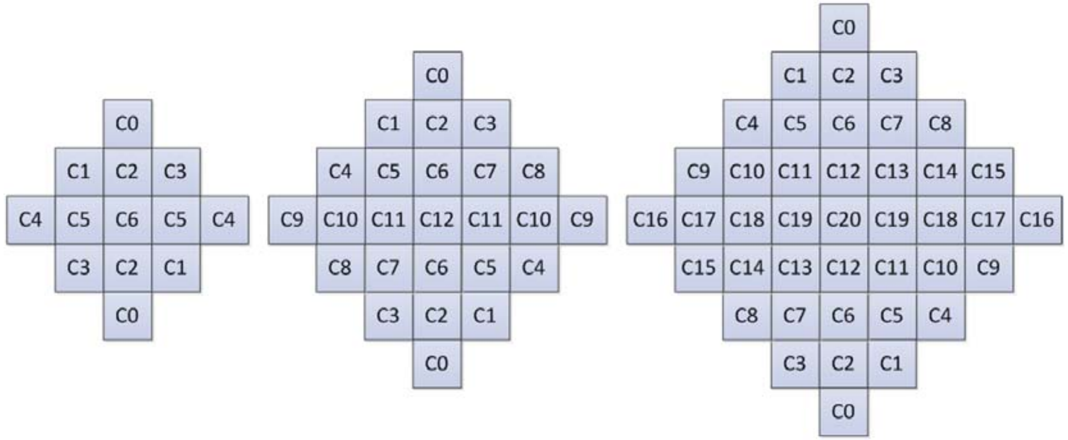




Figures from: JVET-G1001: *Algorithm Description of Joint Exploration Test Model 7*.

JEM: Geometry transform based adaptive loop filter (GALF)

- **Luma component**
  - 25 filters available for each 2×2 block, based on direction and activity of local gradients
  - Diamond filter shapes ($3 \times 3, 5 \times 5, 7 \times 7$)
  - Classification into 25 classes, based on
    - Activitiy index
    - Directionality index
- **Chroma components**
  - Diamond filter shape 5×5
  - No classification
  - Single set of filter coefficients



- Geometric transformations based on data from classification
  - Transpose, vertical flip, rotation

- Filter coefficients signaled with 1st CTU, FIFO buffering for temporal prediction in inter pictures, 16 candidate sets for intra pictures

JEM status / performance

- By mid of 2017, JEM had reached substantial improvement over HM
  - roughly 30% average bit rate reduction for inter coding (random access configuration)
  - roughly 20% for intra-only coding (without motion comp.)
- Subjective tests unveiled that this may even translate into higher visual gains (see JVET-G1004)
  - Mission accomplished: No further development performed afterwards (except bug fixing)
  - JEM was used as additional (well understood) anchor in Call for Proposals ...
  - ... waiting for things to come ...

What was proposed in CfP?

- 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
  - Some new elements for quantization / context-adaptive arithmetic coding

What was proposed in CfP?

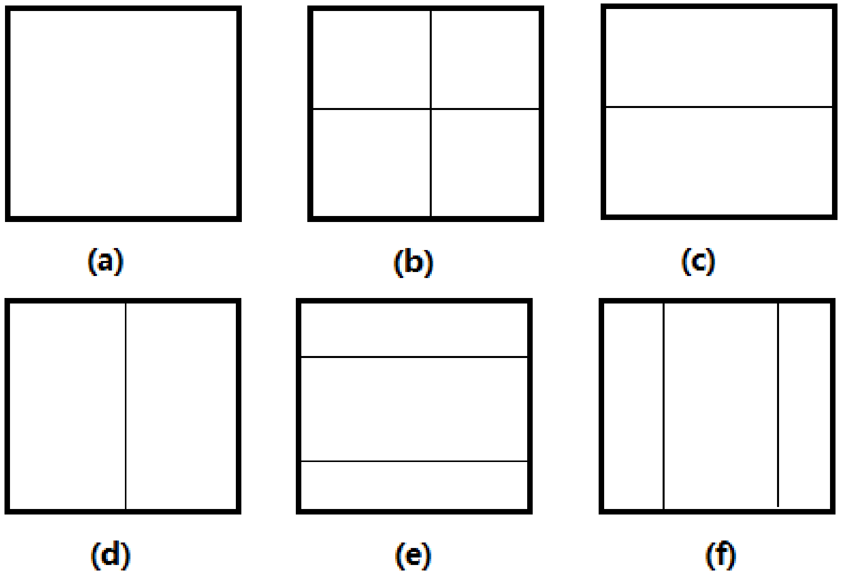
- Compression-improving tools:
  - Template matching tools (decoder side) for purposes of mode/MV derivation and sample prediction both in intra and inter coding
  - Finer partitioning: Asymmetric rectangular, geometric/wedge
  - Enlarged intra reference area & intra block copy
  - Additional non-linear,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

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

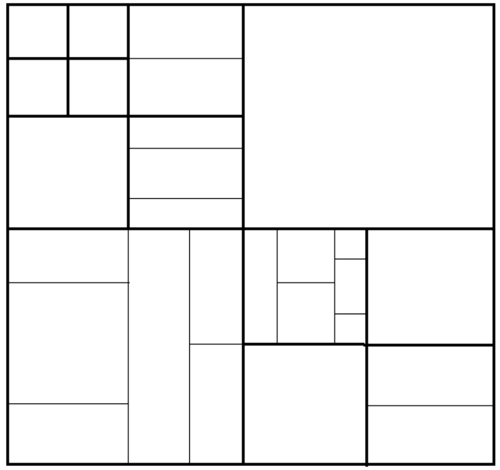
- 360° video specific:
  - Variants of projection formats, geometry-corrected face boundary padding
  - Modification and disabling of existing tools at face boundaries

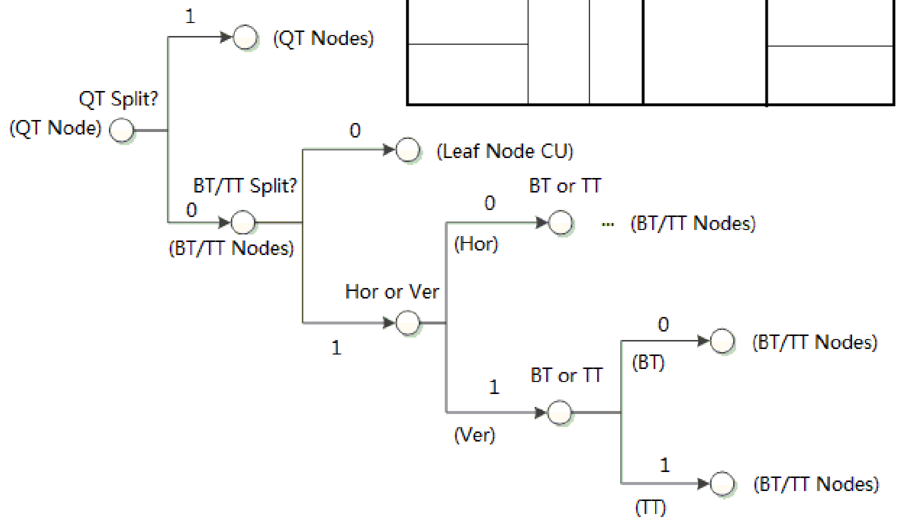
What was new in proposals?

- Simple multi-type tree split was used in several proposals, can be alternated 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

4. Next steps, summary and outlook

Versatile Video Coding – towards the next generation of video compression

PCS 2018 – "Bridging the Gap" Invited Talk

Jens-Rainer Ohm and Gary Sullivan

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

Performance of VTM and BMS compared to HEVC

- PSNR-based CTC BD-Rate savings relative to HEVC reference software (10 bit)

vs HM16.18	VTM	BMS
4k UHD	10%	28%
1080p	8%	22%
WVGA	6%	19%
Average	8%	23%
Decode time	0.8×	2×
Encode time	2×	9×

Documents issued after CfP Results

- Report of Results from the Call for Proposals on Video Compression with Capability beyond HEVC (JVET-J1003)
  - Documentation of results per sequence, marking HM and JEM anchors, not identifying individual proponents
  - Assessment of qualitative (and as far as possible quantitative) benefit of submitted technology compared to anchors
- Working Draft 1 of Versatile Video Coding (JVET-J1001)
  - "Reduced" HEVC plus quad/binary/ternary tree structure
- Test Model 1 of Versatile Video Coding (VTM 1) (JVET-J1002)
  - Corresponding encoder and algorithm description

Core Experiments defined by JVET

- CE1: Partitioning
- CE2: In-loop filters
- CE3: Intra prediction and mode coding
- CE4: Inter prediction and MV coding
- CE5: Arithmetic coding engine
- CE6: Transforms and transform signalling
- CE7: Quantization and coefficient coding
- CE8: Current picture referencing
- CE9: Decoder side MV derivation
- CE10: Combined and multi-hypothesis prediction
- CE11: Composite reference pictures
- CE12: Mapping for HDR content
- CE13: Projection formats

Summary and Outlook

- Video is a lively area of research, major and ongoing progress in standardization
- Though 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
  - Roughly 50% bit rate reduction with same subjective quality as HEVC can probably be reached
  - Rigid process (Core Experiments) just started to establish a reasonable tool combination under complexity/performance/other-acceptability constraints
  - Additional benefit may come from other emerging technology, e.g. deep learning / CNN – if they pass the criteria of bullet points above

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/