

Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team

**Mile High Video Workshop, Denver
July 31, 2018**

Gary J. Sullivan, JVET co-chair

Acknowledgement: Presentation prepared with Jens-Rainer Ohm and Mathias Wien, Institute of Communication Engineering, RWTH Aachen University

1. Introduction

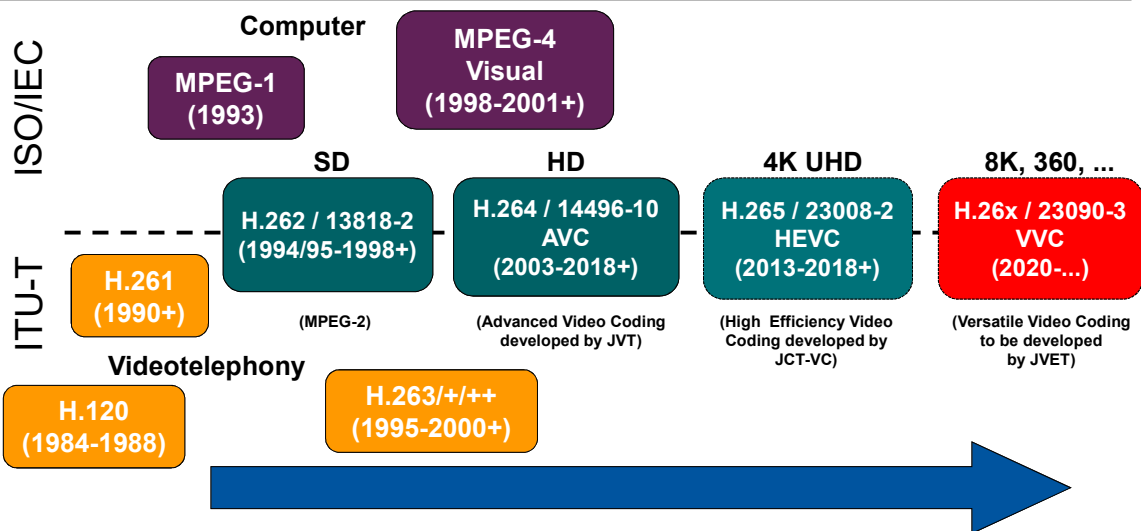
Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team

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 Experts Team”** responsible for developing VVC (established Oct. 2015) – previously called “Joint Video Exploration Team”

3 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

History of international video coding standardization (1985 ~ 2020)



4 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Basic reminders: The motivation for further 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
-

5 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

2. Preparations and Call for Proposals on Versatile Video Coding

Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team

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 (starting Oct. 2015)
 - 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
-

7 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Performance

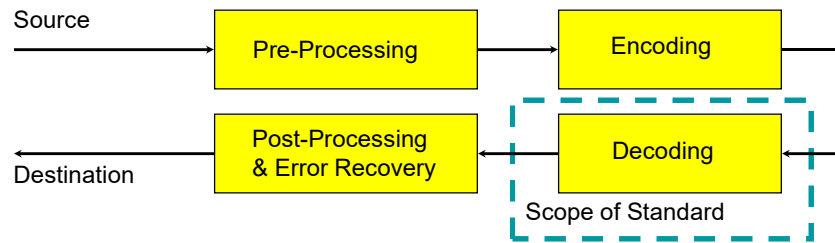
- **Submissions had to provide coded/decoded sequences**
 - 4 rate points each, two constraint conditions "low delay" (LD) and "random access" (RA)
 - **SDR**: 5x **HD** (both LD and RA), 5x **UHD-4K** (only RA)
 - **HDR**: 5x **HD** (PQ grading), 3x **UHD-4K** (HLG grading)
 - **360°** : 5 sequences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

8 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

The scope of video standardization

- Only Specifications of the *Bitstream, Syntax, and Decoder* are standardized:

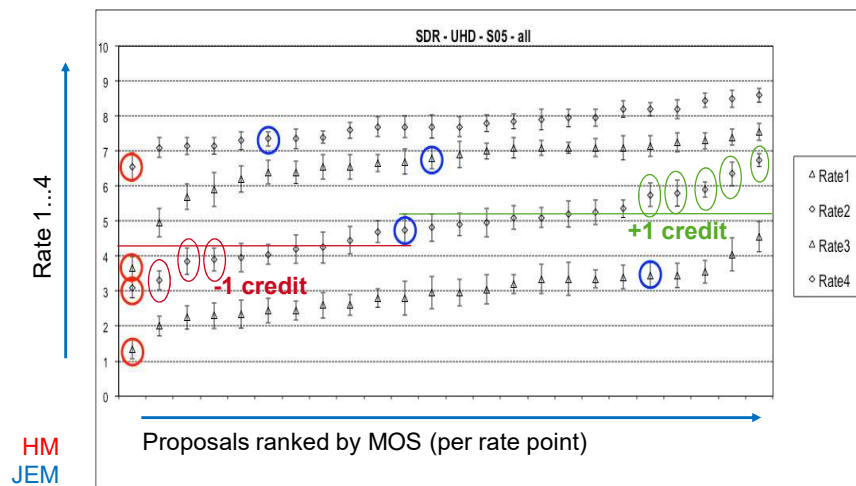
- Permits optimization beyond the obvious
- Permits complexity reduction for implementability
- Provides *no* guarantees of quality



9 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

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



10 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

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
-

11 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
 Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Performance compared to HEVC

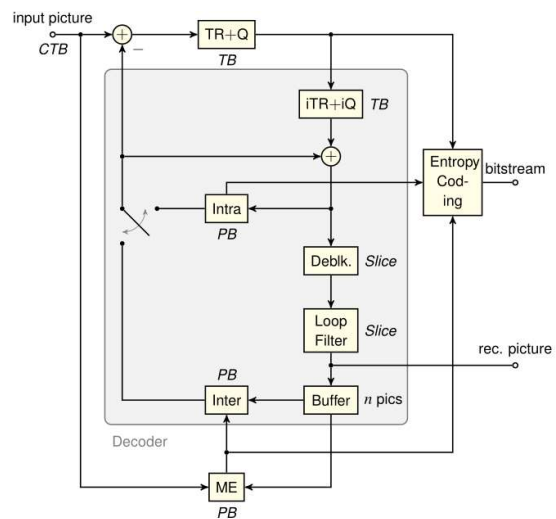
- The subjective quality of best performing proposals is *always equal or sometimes better* (~1/3 of cases) than HEVC at next higher rate point, over all categories (with approx. 40% less rate)
 - The subjective quality of best performing proposals is always equal or sometimes better (~1/5 of cases) 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
-

12 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
 Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

3. Technologies

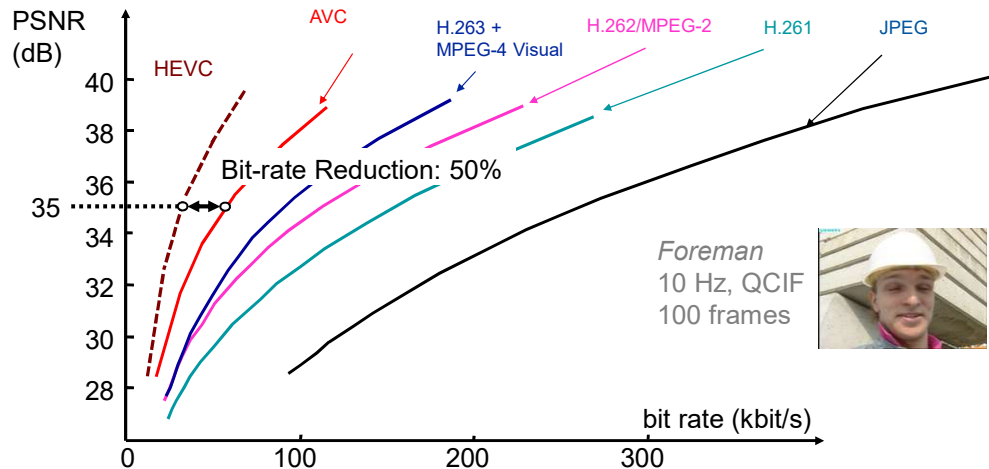
Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team

Hybrid Coding Concept



Basis of every standard since H.261

Performance history of standard generations



15 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

How has video coding been changing?

- **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
 - More sophisticated quantizer optimization
- **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
- **A new twist: Some neural network techniques arising in various elements**

16 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

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
 - Partitioning: 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
 - ...
-

17 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

4. Actions thus far

Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team

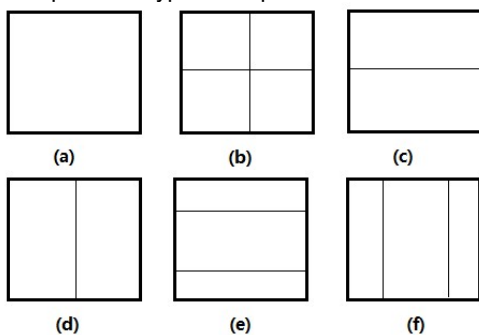
VVC Test Model and Benchmark Set

- **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 1** 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, investigating less mature schemes

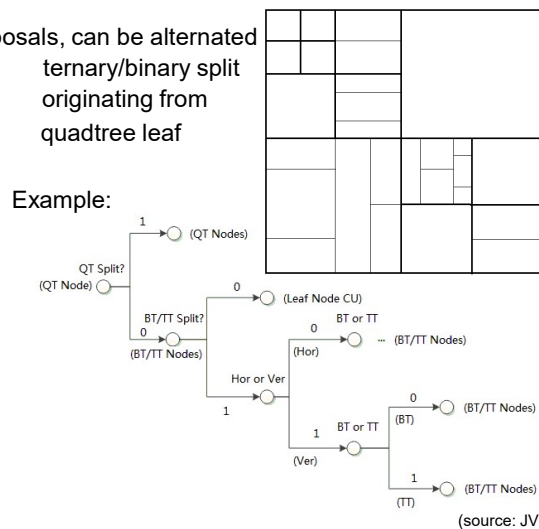
19 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Quadtree with ternary and binary tree

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



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



20 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Performance of VTM 1 and BMS 1 compared to HEVC

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

vs HM16.18	VTM 1	BMS 1
4k UHD	10%	28%
1080p	8%	22%
WVGA	6%	19%
Average	8%	23%
Decode time	0.8×	2×
Encode time	2×	9×

21 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Performance of VTM 2 and BMS 2 compared to HEVC (this month - rough estimation)

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

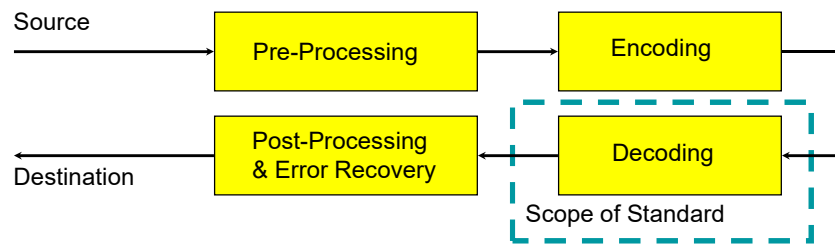
vs HM16.18	VTM 2	BMS 2
4k UHD	~22%?	~30%?
1080p	~20%?	~24%?
WVGA	~17%?	~22%?
Average	~20%?	~25%?
Decode time	~1.8×?	~4×?
Encode time	~4×?	~12×?

22 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

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



23 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Latest status (this month): New elements of WD2 / VTM2

- Remove unnecessary partitioning restrictions
- Implicit splitting at picture boundaries
- Separate trees for intra slices
- Position Dependent Prediction Combination
- Cross Component Linear Model
- 87 intra modes (wide angles included), 3 MPM, TU binarization
- Affine MC (4x4 fixed subblock size, 4/6 parameter model switching at CU level)
- Affine MV coding
 - list construction contains inheritance and derivation spatial/temporal
 - improved difference coding
- Adaptive motion vector resolution (AMVR)
- Subblock MC (4x4) from ATMVP merge, 8x8 granularity motion vector storage [High precision]

24 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Latest status (this month): New elements of WD2 / VTM2

- Multiple transform selection (all are DCT/DST types) for intra and inter
- Increase max QP from 51 to 63
- Modified entropy coding supporting dependent quantization
- Sign data hiding reinvoked from HEVC
- Adaptive loop filter
 - 4x4 classification based (gradient strength & orientation) for luma
 - 7x7 luma, 5x5 chroma filters
 - Enabling flag at CTU level
- Basic high-level syntax (SPS, PPS, slice)
- Update of BMS contains
 - Generalized bi prediction (kind of local weighted prediction)
 - Decoder-side estimation: BIO, simplified bilateral matching
 - Current picture referencing (aka intra block copy)

25 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

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 $T_i(j)$, $i, j=0, 1, \dots, N-1$
DCT-II	$T_i(j) = \omega_0 \cdot \omega_i \cdot \sqrt{\frac{2}{N}} \cdot \cos\left(\frac{\pi \cdot i \cdot (2j + 1)}{2N}\right)$ 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	$T_i(j) = \sqrt{\frac{4}{2N+1}} \cdot \sin\left(\frac{\pi \cdot (2i+1) \cdot (j+1)}{2N+1}\right)$

26 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Cross-Component Linear Model Prediction (CCLM)

- Chroma samples predicted using corresponding reconstructed luma samples

$$pred_c(i, j) = \alpha \cdot rec_L'(i, j) + \beta$$

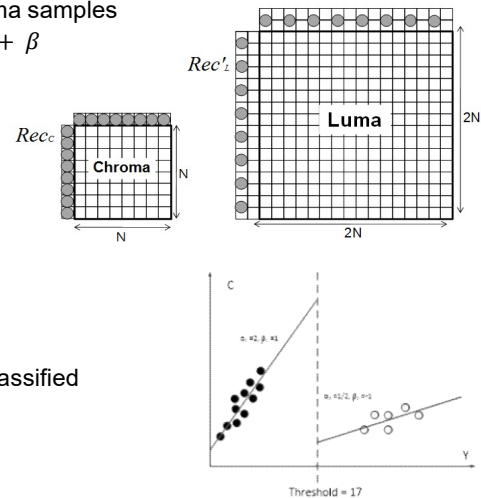
- Parameters α and β : minimize regression error between neighbouring reconstructed luma and chroma samples around current block

- 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.

27 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Extended 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

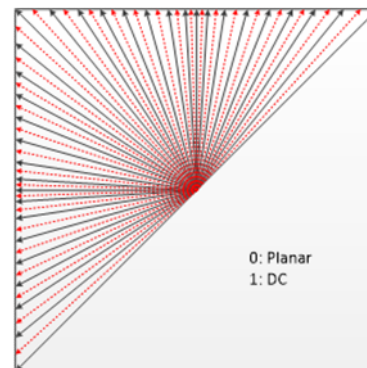
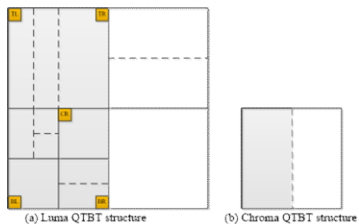
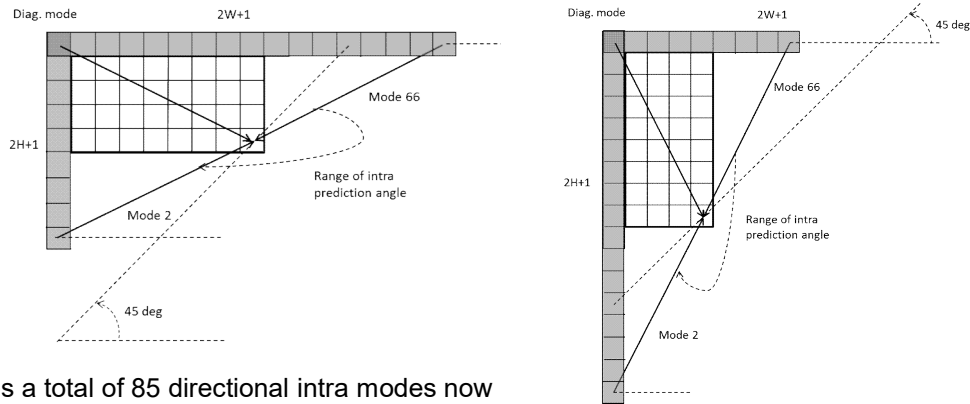


Figure from: Jianle Chen et al. Algorithm Description of Joint Exploration Test Model 7. Doc. JVET-G1001. Torino, IT, 7th meeting: Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Jul. 2017.

28 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Wide angular modes

- For rectangular blocks, prediction directions with angles beyond 45/135 degrees are reasonable
- This can be implemented by adding modes at both ends



- VTM2 uses a total of 85 directional intra modes now (plus DC and planar)

Figures from JVET-K0500

29 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

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)

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

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

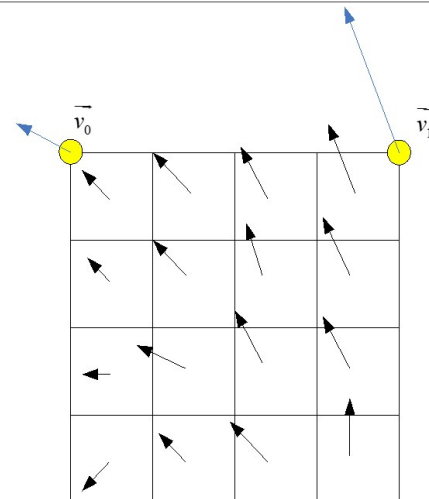


Figure from: JVET-G1001: Algorithm Description of Joint Exploration Test Model 7.

30 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

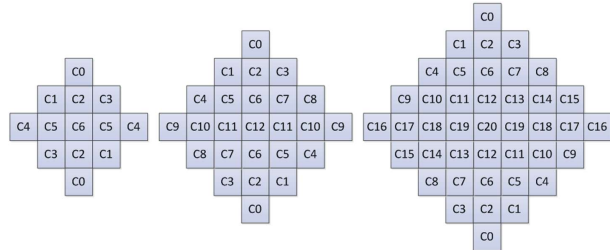
Geometry transform based adaptive loop filter (GALF)

Luma component

- 25 filters available for each 2×2 block, based on direction and activity of local gradients
- Diamond filter shapes (3×3 , 5×5 , 7×7)
- Classification into 25 classes, based on
 - Activity index
 - Directionality index

Chroma components

- Diamond filter shape 5×5
- No classification
- Single set of filter coefficients



Geometric transformations based on data from classification

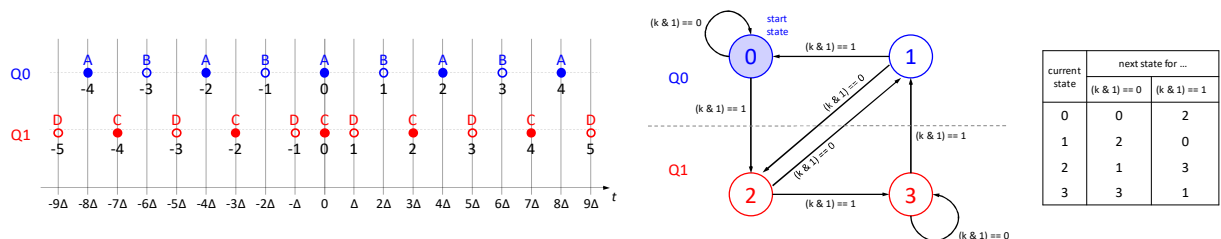
- Transpose, vertical flip, rotation

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

31 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Dependent quantization

- Alternating between two quantizers based on state transition rule allows to select an optimum sequence of reconstruction values (e.g. by trellis-like search)
- Decoder needs to implement the sequential state transition rule
- CABAC contexts needs to be modified as well for this case (greater than 0/1/2/... would have different meaning depending on Q0/Q1)



Figures from JVET-K0071

32 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

Current Core Experiments

- CE1: Partitioning
 - CE2: Adaptive loop filter
 - 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: Deblocking
 - CE12: Mapping for HDR content
 - CE13: Coding tools for omnidirectional video
 - CE14: Post-reconstruction filtering
 - CE15: Palette mode
-

33 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

5. Summary and Outlook

- Video is a lively area of research, major and ongoing progress in standardization
 - 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
 - Formal process (core experiments) in place 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
-

34 Versatile Video Coding – The Next-Generation Video Standard of the Joint Video Experts Team
Gary Sullivan | Jens-Rainer Ohm | Mathias Wien | July 31, 2018

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/