JCTVC-A107

Video coding technology proposal for MPEG-HVC/ITU-T EPVC

K.Sugimoto, Y.Itani, Y.Isu, N.Hiwasa, S.Sekiguchi
Mitsubishi Electric Corporation

R.A.Cohen
Mitsubishi Electric Research Laboratories

P.Wu and N.Sprljan
Mitsubishi Electric R&D Centre Europe, Visual and Sensing Division

Outline

- Codec overview
- Proposed coding tools
- Coding performance
- Complexity analysis
- Simulation software
- Conclusion

Overview of Proposed Codec

- ***MB-based MC+Transform hybrid coding*** with
  - **MB size extension**
  - **Adaptive transform with block-size extension and directional transforms**
  - **Block-based Pyramid Prediction**
  - **Multi-level hierarchical motion partitions including simplified diagonal partition modes**
  - **MB-adaptive Weighted Prediction**
  - **Adaptive PMV/Direct MV derivation**
  - **Adaptive in-loop Wiener filter with extended AVC de-blocking**
- Non-normative remarks to obtain reported performance
  - Single-pass RDO-Q
  - EPZS based motion vector search
  - *No use of frame multi-pass coding*

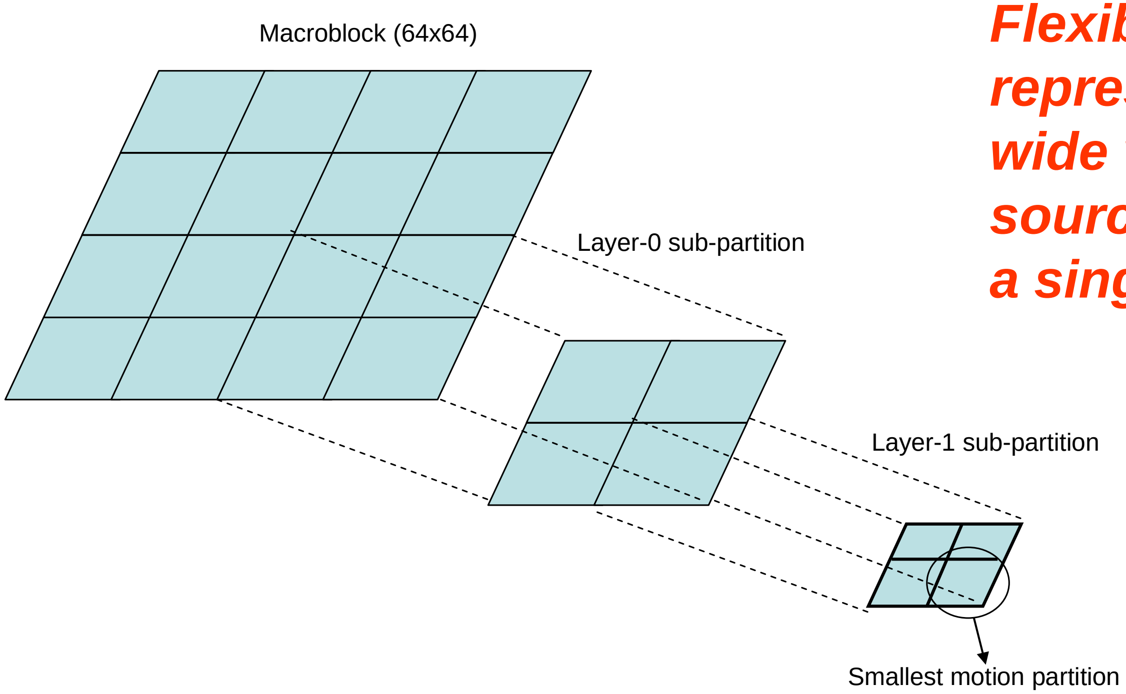
MB size extension

- Promising for coding performance enhancement
  - Better R-D trade-off depending on source characteristics
    - resolution, motion activity, texture smoothness, etc
  - Relevant for video source having higher inter-sample correlation such as ***high-resolution video***
  - Essentially no complexity impact to normative process

Changes for the Better

Motion representation

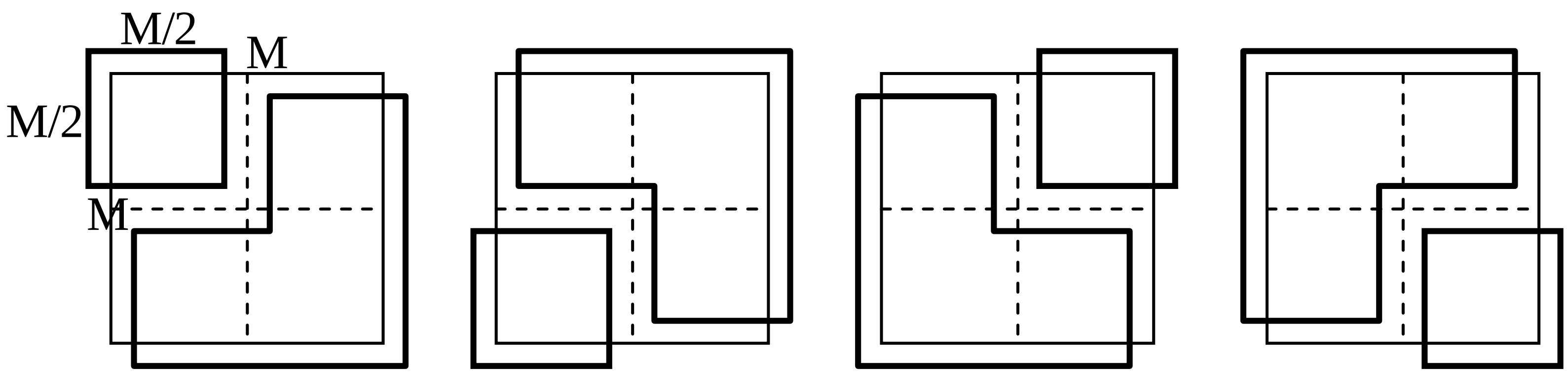
- Multi-layer hierarchical motion partitions
  - Quad-tree based
  - Simplified non-rectangular (“diagonal”) shapes
  - Mixed intra/inter in a MB
  - Hierarchical skip support



Flexible motion representation for wide variety of video source classes with a single syntax

Diagonal motion partitioning

- Promising through literatures
  - Around 8% gain with pixel-wise partitioning (by too-complex encoding)
  - Decoding complexity impact: only on memory access
- Issue: "Efficiency and Complexity" trade-off
  - Flexible partitioning increases pattern representation bits and encoding complexity
- Proposal: ***"Simplified" and "Block-aligned" shapes***
  - PMV derivation by shifting spatial neighboring positions depending on shape

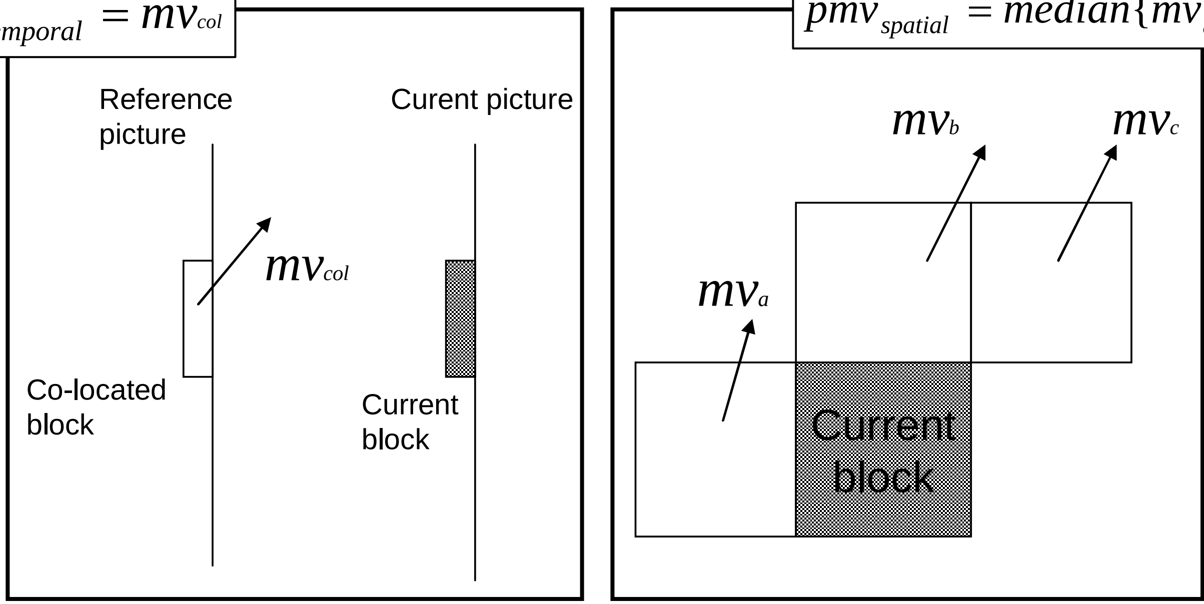


Adaptive PMV derivation

- Extends KTA's Motion Vector Competition concept to exploit temporal correlation
- 1-bit signaling per motion partition to select spatial or temporal PMV candidate

$$pmv_{temporal} = mv_{col}$$

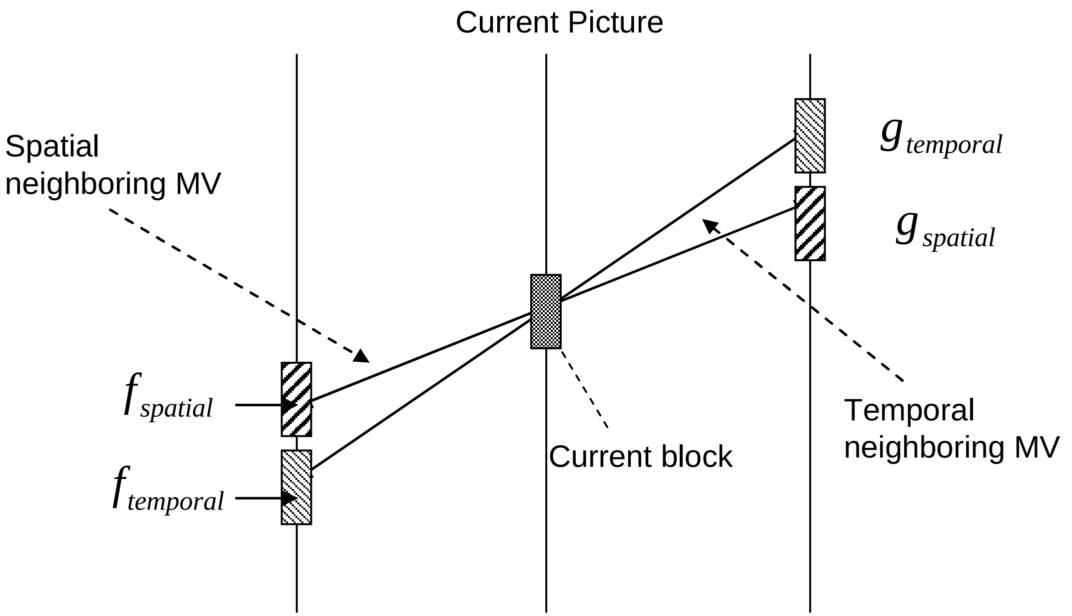
$$pmv_{spatial} = median\{mv_a, mv_b, mv_c\}$$



Contribute to reduction of motion bits especially becoming a burden in high-resolution video coding

Improved direct mode

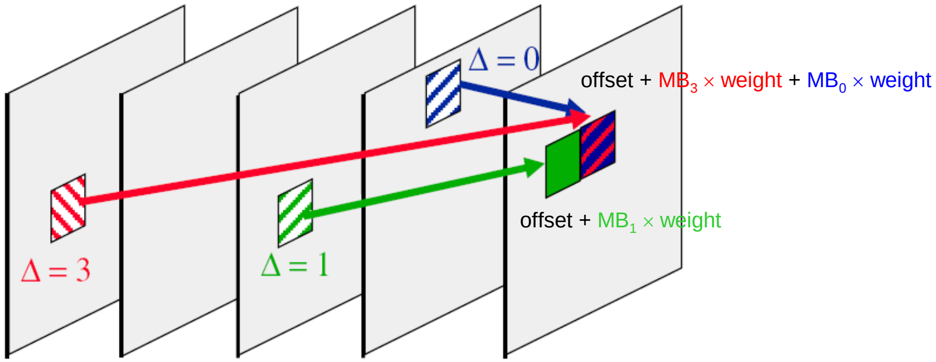
- Adaptive direct MV derivation from spatial and temporal candidates
  - Without signaling bit by SAD competition at both encoder and decoder
  - Used also as B-Skip MV
- Use of closest spatial neighboring MV, not outside a MB as AVC
  - Consideration on MB size extension
- Selective prediction generation at picture edge
  - “Uni-directional pred.” when UMV observed at picture edge
  - Avoid unreliable prediction with UMV



Contribute to coding performance improvement for B-picture

MB-adaptive Weighted Prediction

- WP in H.264 / MPEG-4 AVC works at a picture level - $\mathbf{Y} = o + w\mathbf{X}$
- Explicit mode parameters computed for the picture prior to compression, commonly used in combination with Picture Level RDO
- MBWP defines WP signalling and operation at the MB level - $\mathbf{MBp} = o + w\mathbf{MBc}'$



- Decoder: Complexity as in AVC-style WP.

- Encoder: Lagrangian cost minimisation with MV and WP (weight and offset) as variables. Minimum distortion point can for instance obtained by linear regression and then the result is modified by a rate-constrained search:

$$J(o,w,mv) = D(o,w,mv) + \lambda_{WP} R(o,w) + \lambda_{MV} R(mv)$$

Other aspects on inter prediction

- Essentially identical to AVC
  - Semantics of slice (picture) type
  - Q-pel accuracy MV representation
  - Chroma MV derivation
  - Multiple reference pictures and refIdx signaling
  - MC interpolation filtering

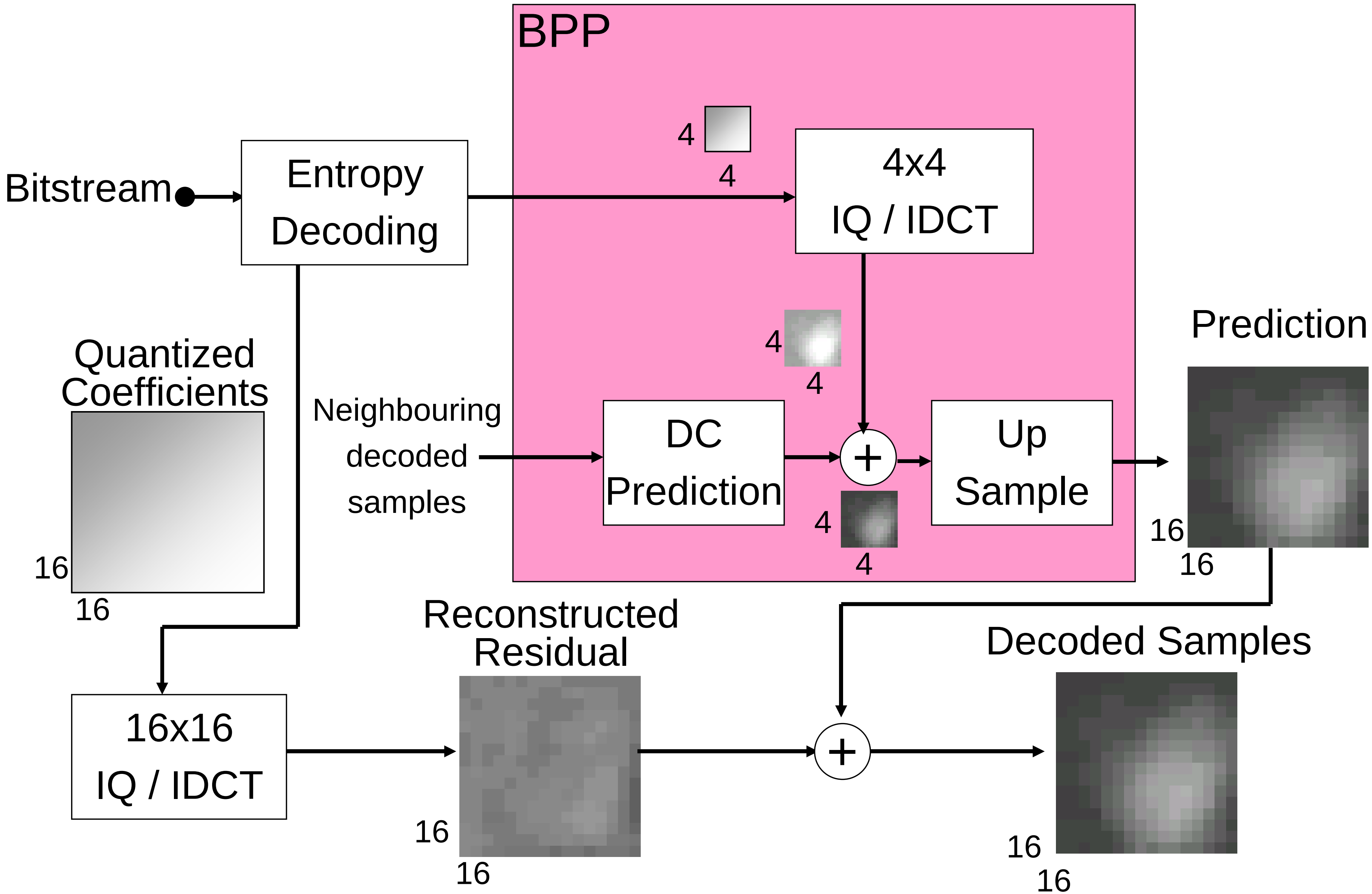
Intra prediction

- New intra coding mode
  - **Block-based Pyramid Prediction (BPP)**
- Re-use AVC on
  - Luma intra4x4, intra8x8, intra16x16 in addition to BPP
- Modification to AVC on
  - 16x16 DCT for intra16x16 residue
  - Intra8x8 coding to chroma

Block-based Pyramid Prediction (BPP)

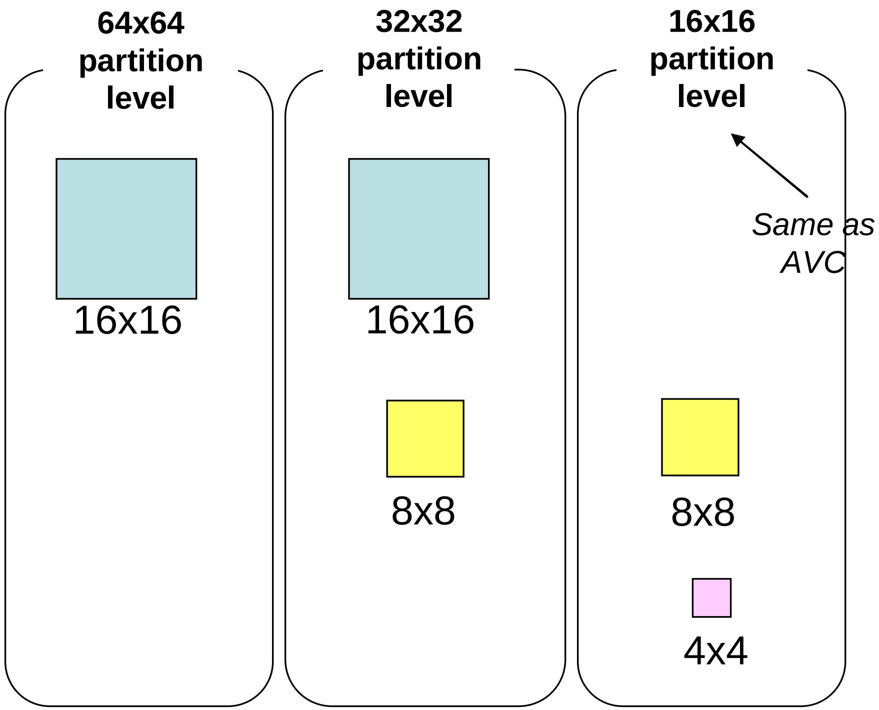
- Provides “rough sketch” of the original block samples as the prediction.
  - Better prediction compared with the intra prediction in AVC could be provided for natural images such as non-straight edges, spot, etc.
- Hybrid BPP and AVC intra prediction algorithm
  - IntraNxN (N=4,8,16) predictions in AVC are also supported.
  - RDO decision of the best intra prediction
- Could easily be extended to support larger block size, 4:4:4 or 4:2:2 video formats.

BPP Normative Process



Adaptive 2-D Transform

- Adaptive block-size 2D transform for luma inter residue
  - **Simple decision** depending on motion partition size

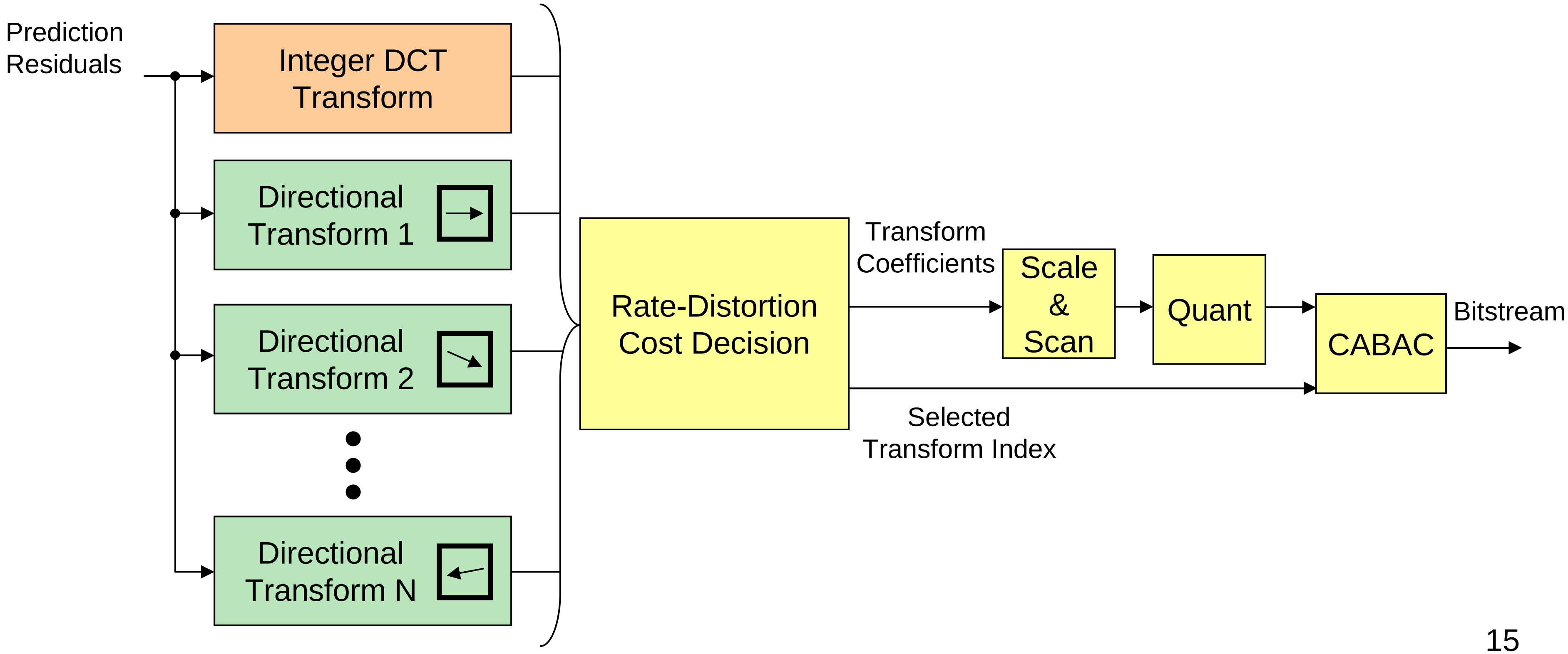


Efficient signal power concentration depending on residual statistics

- Coding mode dependent transform for luma intra residue
  - intra4x4, intra8x8 → 4x4, 8x8
  - Intra16x16 → 16x16 DCT

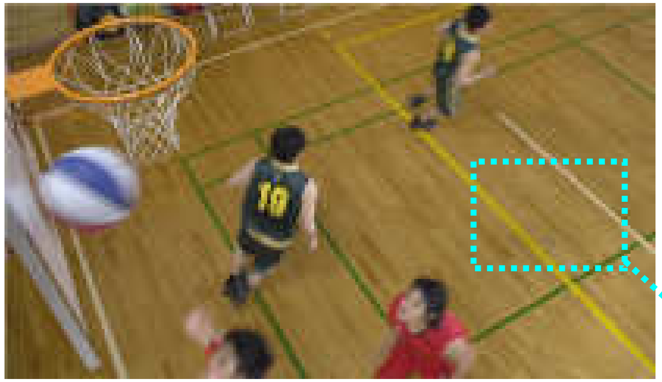
Directional Transforms

- Directional transforms used in addition to 2-D DCT
- Sets of low-complexity 1-D transforms co-aligned over block
- Conventional transforms are used – no training required
- Rate-distortion optimized selection of transform

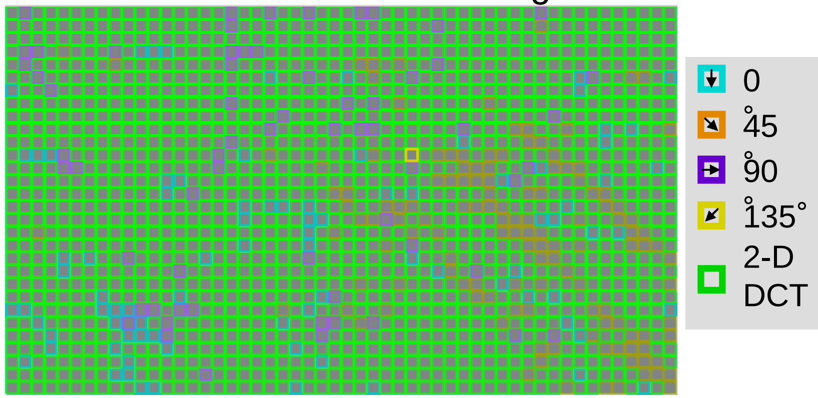


Example Using Directional Transforms

Original BasketballDrill



Directional Transform Usage



Intra Picture Coded without Directional Transforms

Intra Picture Coded with Directional Transforms

Quantization/Scanning

- Conceptually identical to AVC
  - 16x16 DCT does not use decomposition into “core” + “scale” matrices, and zig-zag scanning is used
- Modification for Directional Transforms
  - Coefficients scaled so existing quantizer can be used
  - DC coefficients scanned first, then AC
- Non-normative settings for obtaining reported coding performance
  - Single-pass RDO-Q **USED**
  - Hierarchical QP offsets **USED**
  - Q-matrix function **NOT USED**

Entropy Coding

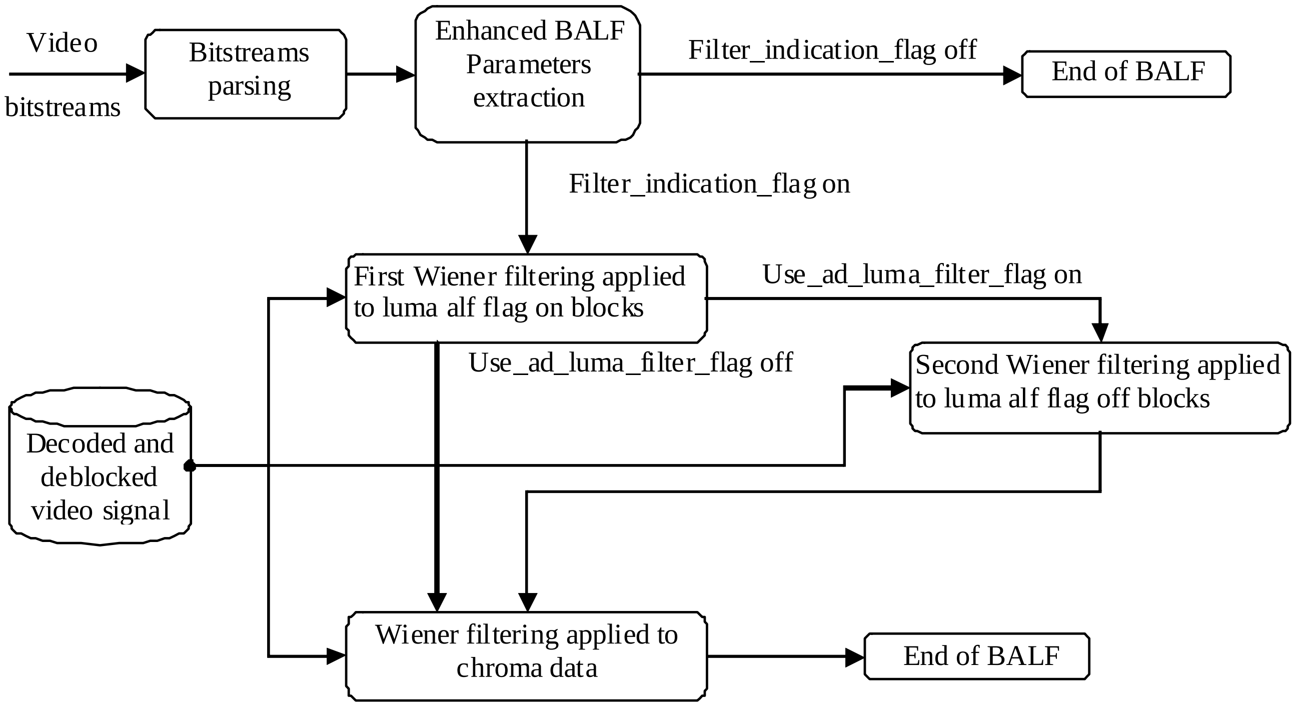
- **CABAC** with modifications to support MB size extension syntax
  - Quad-tree based CBP representation
  - Unification of symbol representation of intra16x16 mode with other intra modes
  - Skip mode at sub-partition levels

In-loop filtering

- Extended AVC de-blocking filter
  - Consideration on 16x16 transform block and mixed intra/inter blocks
  - Strength derivation and actual filtering process identical to AVC
- Adaptive Wiener filter
  - Enhancement to KTA BALF tool
  - Quality improvement of decoded picture and inter-prediction reference

Enhanced BALF

- Applies up to two Wiener filters per video component adaptively
  - Optimising two Wiener filters together with the adaptive indication flags



<u>Significant coding gain</u> observed,
at penalty of picture-level repetitive filter design
and local adaptation process

Coding Performance Summary

- Constraint Set1
  - Hierarchical B structure identical to Alpha anchor
  - Around ***1dB luma BD-PSNR gain for class A/B/C*** at higher compression condition
- Constraint Set2
  - IPPP structure with hierarchical P style QP offsets
  - Relative to Beta anchor
    - Around ***1dB luma BD-PSNR gain for class B/E*** at higher compression condition
  - Relative to Gamma anchor
    - Around ***2dB luma BD-PSNR gain for class B/E*** at higher compression condition

Class A/B, CS1 vs. Alpha anchor

Sequence	BD low		BD high	
	Bitrate Δ	PSNR Δ	Bitrate Δ	PSNR Δ
Traffic	-25.14	1.07	-22.41	0.83
PeopleOnStreet	-16.88	0.95	-13.86	0.76
avarage	**-21.01**	**1.01**	**-18.13**	**0.79**

Sequence	BD low		BD high	
	Bitrate Δ	PSNR Δ	Bitrate Δ	PSNR Δ
Kimono	-31.81	1.35	-29.50	1.08
ParkScene	-20.90	0.85	-16.70	0.65
Cactus	-24.17	0.85	-22.88	0.68
BasketballDrive	-27.47	1.03	-25.00	0.78
BQTerrace	-33.04	0.69	-32.80	0.53
avarage	**-27.48**	**0.95**	**-25.38**	**0.74**

Class B/E, CS2 vs. Beta anchor

Sequence	BD low		BD high	
	Bitrate Δ	PSNR Δ	Bitrate Δ	PSNR Δ
Kimono	-33.68	1.52	-32.43	1.31
ParkScene	-15.09	0.58	-12.01	0.44
Cactus	-18.20	0.63	-17.26	0.52
BasketballDrive	-29.54	1.18	-27.63	0.95
BQTerrace	-27.32	0.63	-25.05	0.47
avarage	**-24.76**	**0.91**	**-22.87**	**0.74**

Sequence	BD low		BD high	
	Bitrate Δ	PSNR Δ	Bitrate Δ	PSNR Δ
Vidyo1	-27.70	1.38	-26.55	1.10
Vidyo3	-20.05	1.00	-20.85	0.86
Vidyo4	-20.75	0.95	-18.80	0.70
avarage	**-22.83**	**1.11**	**-22.07**	**0.89**

Class B/E, CS2 vs. Gamma anchor

Sequence	BD low		BD high	
	Bitrate Δ	PSNR Δ	Bitrate Δ	PSNR Δ
Kimono	-48.99	2.51	-49.01	2.28
ParkScene	-36.77	1.64	-32.73	1.42
Cactus	-40.23	1.64	-38.70	1.44
BasketballDrive	-45.18	2.13	-43.39	1.76
BQTerrace	-56.02	1.89	-51.29	1.43
avarage	**-45.44**	**1.96**	**-43.02**	**1.66**

Sequence	BD low		BD high	
	Bitrate Δ	PSNR Δ	Bitrate Δ	PSNR Δ
Vidyo1	-44.79	2.58	-44.63	2.16
Vidyo3	-40.14	2.33	-40.08	2.01
Vidyo4	-41.42	2.23	-41.87	1.87
avarage	**-42.12**	**2.38**	**-42.20**	**2.01**

Complexity in Encoding Process

- Similar degree of encoding speed to JM anchor observed
  - Reported performance results were obtained
    - **at around 2x encoding time than anchor with current functional algorithm simulator**
    - **without using frame-level multi-pass coding**
    - **same memory size as anchor**
- Complexity notice
  - Computations for MB-level RDO decision due to increased number of motion partitions and prediction/coding modes
  - Total memory requirement strongly depends on implementation, and can be expected via conventional codec implementation

Mode decision in I-picture

- 16x16 conventional MB size
- All prediction modes and MB coding modes are determined by RDO
- Transform decision
  1. RDO decision with 2-D transform
  2. Directional transform basis functions further evaluated by RDO for intra8x8 and intra16x16 cases
- Chroma
  – intra8x8 coding

Mode decision in P/B-pictures

- MB size: 64x64(class A/B) or 32x32(class C/D/E)
- Motion estimation
  - EPZS based
  - ME cost: sum of SADs of all color components and motion bits
  - MBWP parameter decision
- Mode and transform block-size/basis type
  - jointly determined by RDO per each motion partition
- Intra mode
  - evaluated at partition having 16x16 block size, after finding the best MV and inter mode

Complexity in Decoding (Normative) Process

- Could be within a few times complexity relative to AVC High Profile
- Computational complexity factors
  - In-loop Wiener filtering
  - Down/Up-sampling in BPP (less impact for inter pictures)
  - SAD computation for adaptive direct MV derivation
- Same level of computational complexity as AVC for
  - Directional Transforms
  - MB-adaptive Weighted Prediction
- Additional memory usage
  - Intermediate storage for adaptive Wiener filter process
- Additional memory bandwidth
  - SAD computation for adaptive direct MV derivation
  - Frame memory access during in-loop Wiener filtering

Other characteristics

- Random access & Delay characteristics
  - Same degree as existing MPEG/ITU-T H.26x Standards by using conventional GOP structure
- Error resilience
  - Same degree as AVC in terms of context adaptive coding architecture
  - Slice structure is possible as usual
- Extensibility
  - Supports of 4:4:4 coding, scalability, multi-view coding are equivalently possible as done in AVC extensions
- Parallel processing
  - Conventional approaches possible (Slice partitioning, MB-level pipeline process flow)
  - In-loop Wiener filter requires more consideration

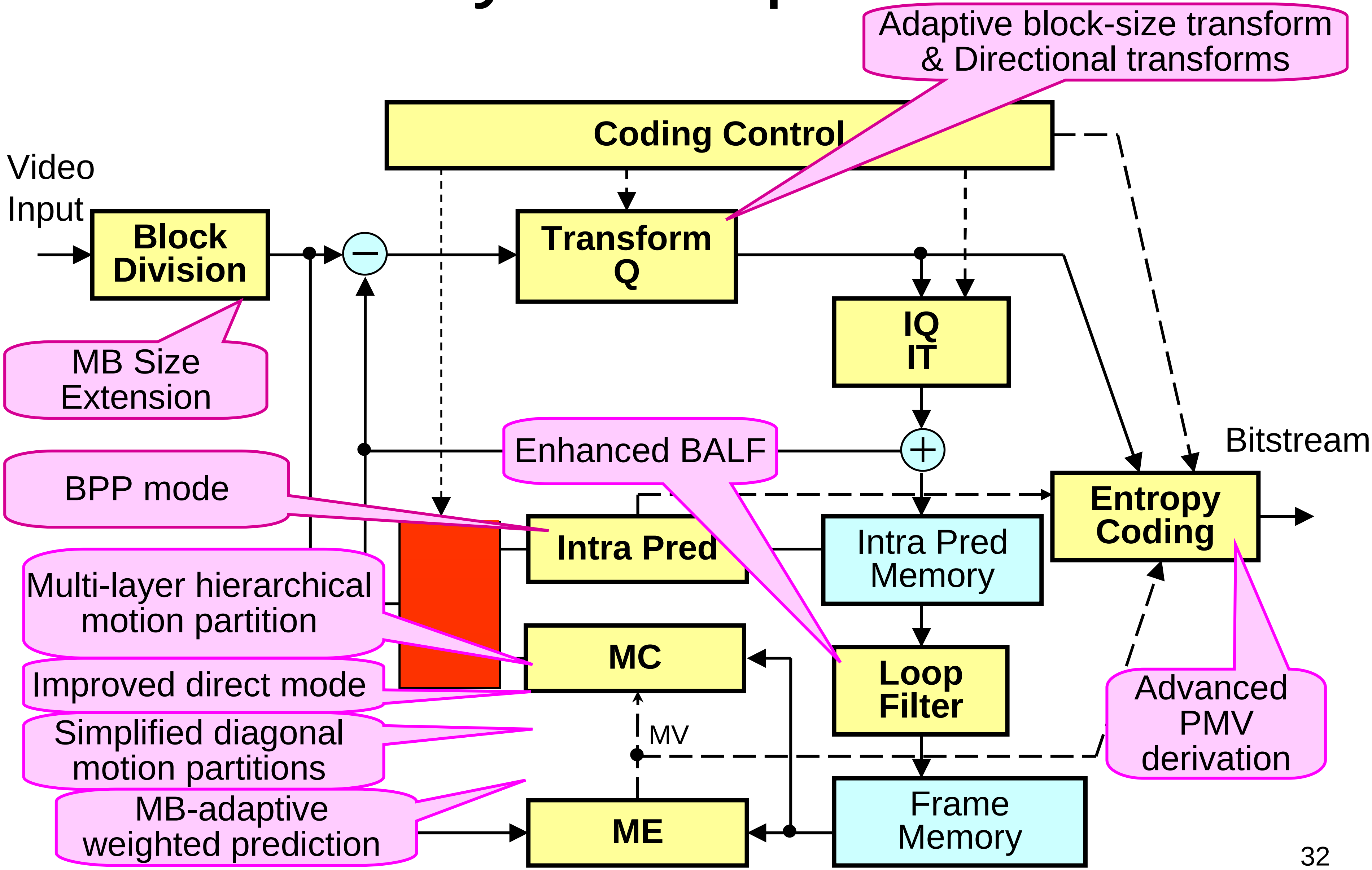
On simulation software

- Pure standard C-language based code
  - Clean implementation of extended MB size support
  - No use of CPU dependent instructions
- Multi-platform
  - Windows (VisualStudio, gcc/cygwin)
  - Linux (gcc)
  - Mac OS X (gcc)
  - Any UNIX workstations (gcc)
- 64bit operating system assumed
- Modularized extensible design
  - existing software modules in JM/KTA has been re-using without changing the code too much

Remarks

- MB-based hybrid coding is still preferable for practical codec implementations
  - Especially for next-generation UHDTV H/W codec up to 8Kx4K
- MB size extension and in-loop Wiener filter tool should be considered as “starting point” of reference model
  - has been verified by explorations in MPEG CfE / VCEG KTA study
  - Significant coding gain achieved
- Core experiments should target:
  - More performance enhancements with additional proposed tools
  - Further studies on performance-complexity trade-off
  - Performance improvement for higher resolution video sources
- Our codec simulator offers clean and extensible platform for future collaborative standardization work

Summary of Proposed Tools



Summary of Proposed Tools

Category	Proposed Tool
MB size	Explicit MB size signaling
Transform	•Adaptive block-size transform •Directional transforms
Entropy Coding	Advanced PMV derivation
Loop Filter	Enhanced BALF
Intra Prediction	Block-based Pyramid Prediction
Motion representation	•Multi-level hierarchical motion partitioning •diagonal partition shapes •Improved direct mode •MB-adaptive weighted prediction

Conclusion

- A new video coding algorithm for MPEG-HVC/ITU-T EPVC presented
  – Single-pass picture coding with “Back-to-basics” approach
  – Up to 1dB Luma PSNR gain relative to Alpha/Beta anchor
  – Up to 2dB Luma PSNR gain relative to Gamma anchor
  – Predictable decoder complexity relative to AVC
- Recommendations
  – To adopt the proposed coding algorithm framework into reference model and enhance performance further through CE process until Oct.
  – To consider Super Hi-Vision source for further studies
    - To be proposed by JCTVC-A023