

Description of SDR, HDR, and 360° video coding technology proposal by InterDigital and Dolby

JVET-J0015

Introduction

- JEM is an experimental platform
 - Provides good coding performance
 - For CS1, JEM anchors provide 33.15% average BD rate savings over HM CfP anchors
 - But has many complexity issues, not just in terms of encoding/decoding time
- Many JEM coding tools are well understood by the experts, and some simplified flavor of them will likely be included in our next standard
- We target simplification with the following important design considerations for the future video coding standard

Decoder complexity is still one of the most important design aspects

Any increase in decoder complexity must be justified

Real-time decoder must be implementable in the near-term

The SDR codec should serve as the core coding engine for HDR and 360 video

Modifications/customizations can be allowed if justified, e.g., subjective quality improvement

Summary of J0015

- This joint response follows these design considerations
 - For SDR:
 - Great attention is paid to reduce decoder complexity, many tools are simplified with little performance loss
 - Further performance improvement achieved through 2 additional tools

	BD rate savings			Enc time *	Dec time
	Y	U	V		
CS1	35.72%	44.75%	44.95%	1710%	263%
CS2	27.18%	43.81%	44.51%	1827%	301%

* Encoding time increase mostly due to MTT implementation

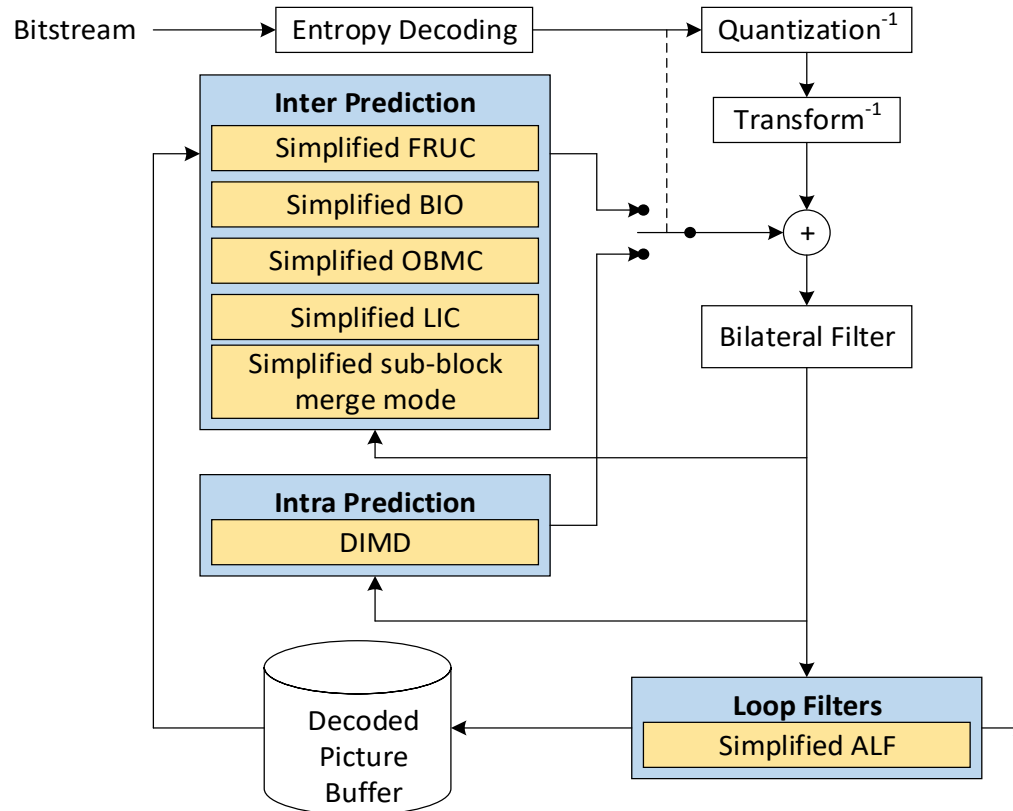
- For HDR:
 - Improve HDR coding efficiency using coding tools that reside inside the core codec
- For 360 video:
 - Main goal is to solve face seam artifacts for this type of video → small modifications are proposed to some SDR tools
 - Exploration of projection format as a “coding tool” for coding performance improvement
- This presentation is organized by categories

SDR Category

SDR key features

- Decoding complexity reduction
 - Simplified frame-rate up conversion (FRUC)
 - Simplified bi-directional optical flow (BIO)
 - Simplified overlapped block motion compensation (OBMC)
 - Simplified local illumination compensation (LIC)
 - Simplified sub-block merge mode
 - Simplified adaptive loop filter (ALF)
- Coding performance improvement
 - Multi-type tree (MTT)
 - Decoder-side intra mode derivation (DIMD)

SDR key features



SDR Category: simplification of existing JEM tools

Simplified frame-rate up conversion (FRUC)

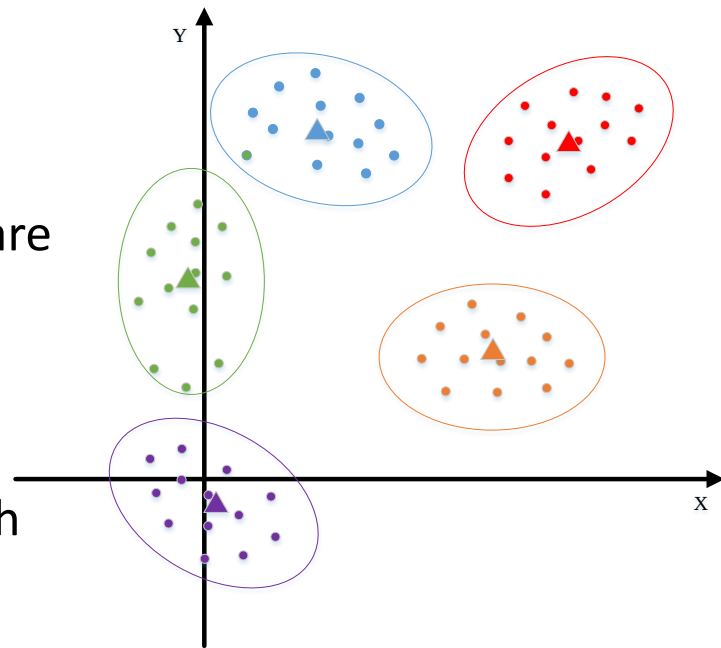
- The FRUC in JEM-7.0 is used as the base
 - Two FRUC modes: template-matching and bilateral-matching
 - Both FRUC modes follow a 4-step process
 - CU-level motion search, CU-level motion refinement, sub-CU-level motion search and sub-CU-level motion refinement
- Proposed FRUC simplifications
 - **MV clustering method** at CU-level and sub-CU-level motion search
 - **Multi-stage motion refinement method** at CU-level and sub-CU-level motion refinement
 - **Early termination** to skip sub-CU-level motion search and sub-CU-level motion refinement
 - **One constrained search range** applied to the entire FRUC process

FRUC: MV clustering based motion search (1/2)

- Initial MV candidates include the following
 - Same as in JEM-7.0
 - AMVP predictors if the CU is coded by AMVP
 - Regular merge candidates
 - Spatial neighboring MVs from the top and left neighbors
 - Different from JEM-7.0
 - The interpolated MV candidates are removed
 - ATMVP and STMVP candidates included only at sub-CU-level motion search
- Reference picture selection
 - Select the reference picture most frequently used by initial MV candidates
 - MV scaling and pruning
 - Scale the MV candidates to the selected reference picture(s)
 - Remove any redundant MV candidates
 - Benefits
 - ***Only one reference picture per L0/L1 list*** need to be accessed during FRUC motion search
 - Reduce the memory bandwidth

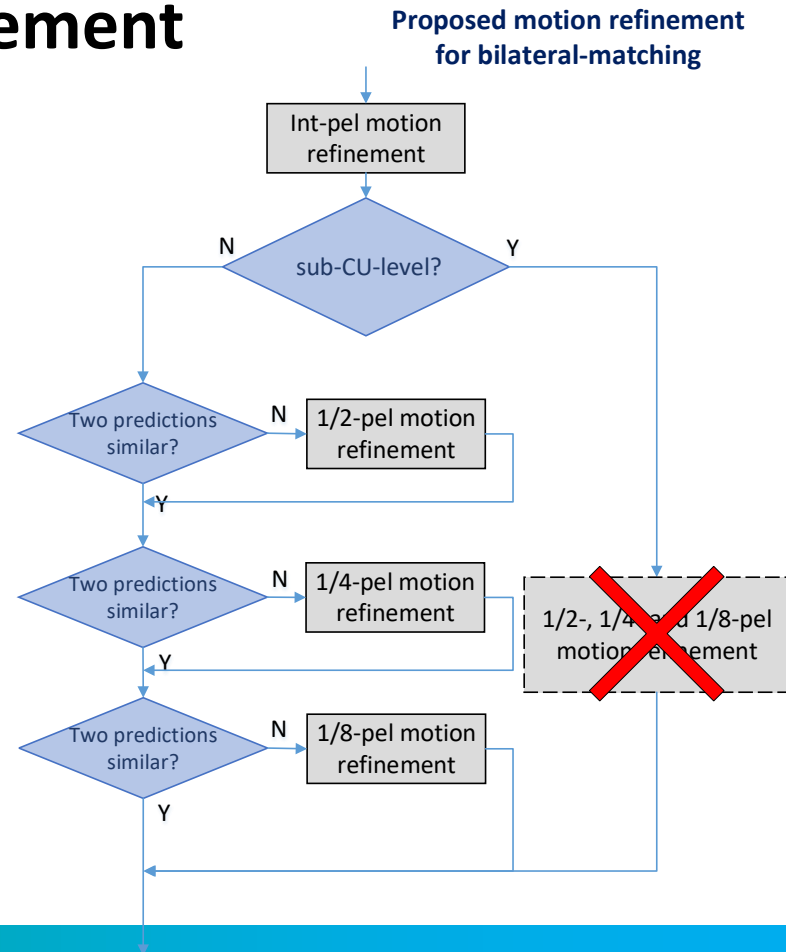
FRUC: MV clustering based motion search (2/2)

- MV clustering
 - Cluster initialization
 - Cluster assignment based on MV similarity
 - Cluster update
 - The centroids of the resulting MV clusters are used for CU-/sub-CU-level motion search
- Advantages
 - Reduce the number of tested MVs
 - The clustering process is performed on each MV candidate only once
 - No iteration is needed, very low complexity



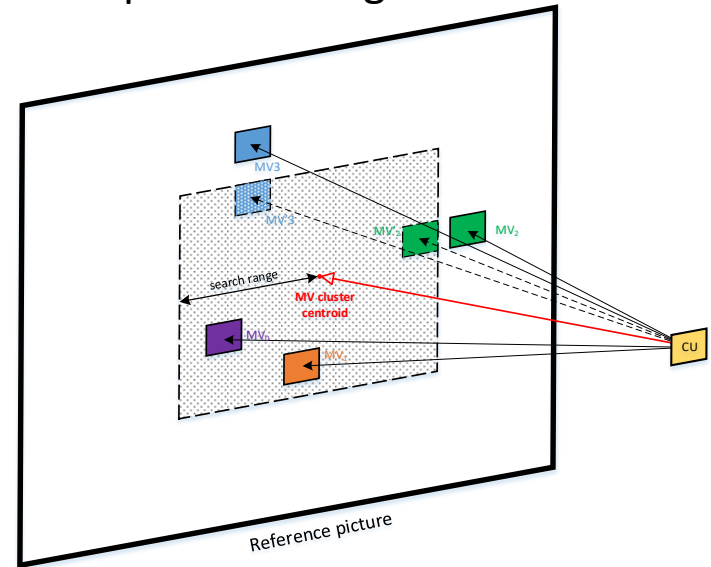
FRUC: Multi-stage motion refinement

- Motion refinement in JEM-7.0
 - MV is searched at 1/4-pel steps
- Proposed multi-stage motion refinement
 - Integer-pel motion refinement
 - Perform integer MV search around the initial MV and update the search center
 - **At most 12 search iterations**
 - Fractional-pel motion refinement
 - Improve the MV precision progressively at 1/2-, 1/4- and 1/8-pel
 - Disable sub-CU-level fractional motion refinement
 - For FRUC bilateral-matching, skip fractional-pel motion refinement based on prediction similarity
 - Template-matching uses a similar multi-stage process
 - Combined with early termination (next slide)



FRUC: Other simplifications

- Early termination of sub-CU-level motion derivation considering the following conditions:
 - If FRUC template-matching
 - If FRUC bilateral-matching and if the two CU-level prediction signals are similar
- Constrained search range for the FRUC
 - One search region of 96×96 luma samples is determined around the centroid of the largest MV cluster
 - Clip each tested MV to be within the search region
 - ***Single memory access to only one reference picture per list is needed***



Simplified Bi-directional optical flow (BIO)

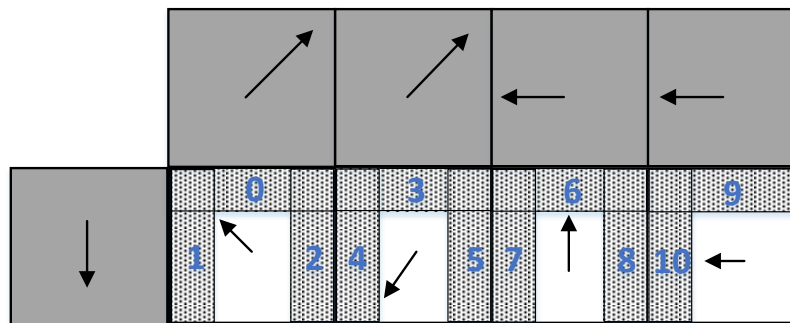
- Two-stage early termination of the BIO
 - Conditionally skip the BIO refinement based on the similarity between two prediction signals at both CU-level and sub-CU-level
- The modified filtering order for the BIO gradient derivation
 - For gradient derivations, horizontal filter is always firstly applied (sample or gradient), followed by vertical filter (gradient or sample)
 - Better aligned with the other filtering processes, such as MC and de-blocking
- The modified rounding operation
 - The BIO in JEM-7.0 uses logical right-shifts
 - Calculate the absolute of input, right-shift the absolute, and multiply the sign
 - Proposed
 - Use arithmetic right-shifts, i.e., the sign is kept during the right-shifts
 - Avoid the need to calculate absolute values

Simplified overlapped block motion compensation (OBMC)

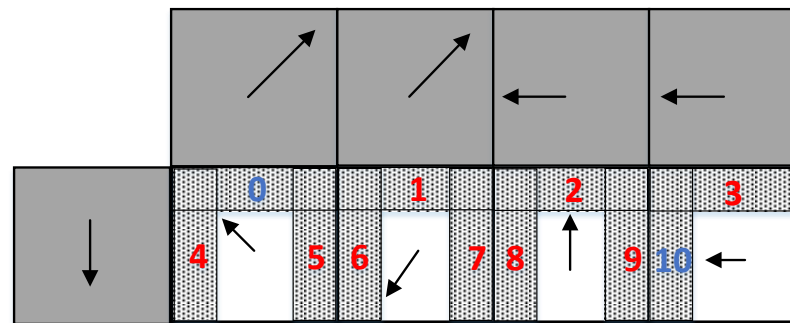
- The OBMC in the JEM-7.0
 - Processing is based on 4x4 block units following raster scan order
 - Only being skipped when neighboring 4x4 blocks have the same motion
- Observations
 - With the existing OBMC order, the MCs of neighboring 4x4 blocks cannot be merged even when they have the same motion
 - OBMC is less effective when neighboring 4x4 blocks have similar motion
- Simplified OBMC
 - **OBMC block grouping** is used to reduce the number of MCs during the OBMC
 - **Early termination** to skip OBMCs when the neighboring blocks have similar motion, i.e., same reference picture(s) and the MV difference is no larger than 1/2-pel

OBMC: Block grouping (1/2)

- Divide the OBMC into two steps
 - **Boundary OBMC:** adjusting the samples on the CU boundaries using the MVs from spatial neighboring CUs
 - **Internal OBMC:** adjusting the samples on the sub-block boundaries, except the samples on the CU boundaries
 - The order of adjusting the boundary samples of the CU is changed



JEM-7.0

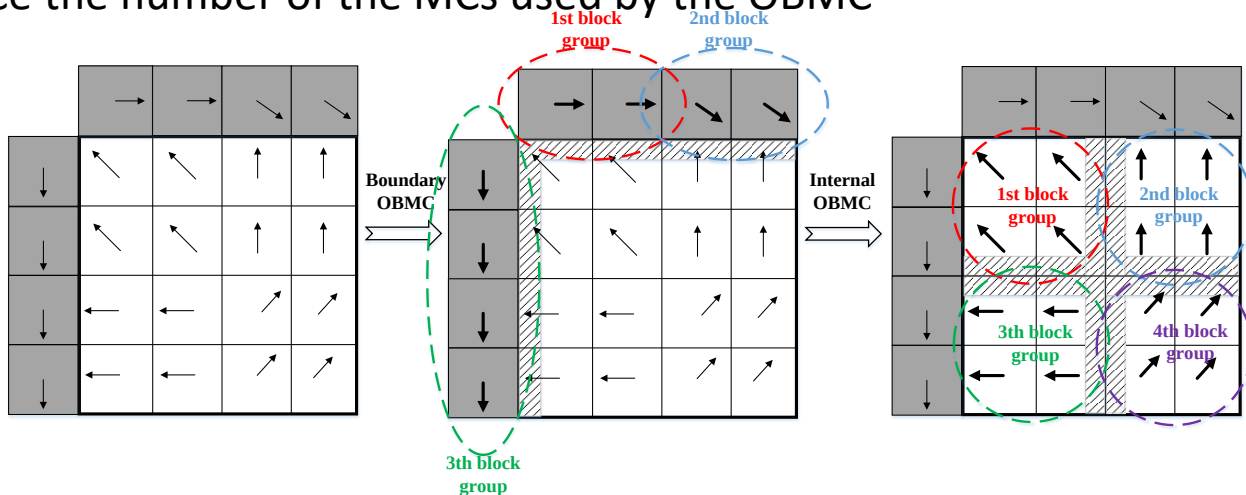


Proposed

OBMC: Block grouping (2/2)

- Block grouping in the OBMC

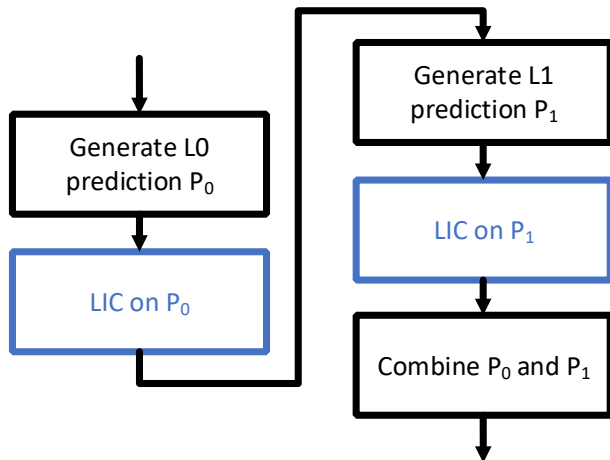
- Boundary OBMC: merging the OBMCs of adjacent neighboring 4x4 blocks on the top/left CU boundary with the same motion into one single MC
- Internal OBMC: aligning the sub-block size used for the OBMC with that used for the sub-block-level motion derivation, e.g., FRUC, ATMVP → larger than 4x4 sub-block size may be used
- Reduce the number of the MCs used by the OBMC



Simplified local illumination compensation (LIC)

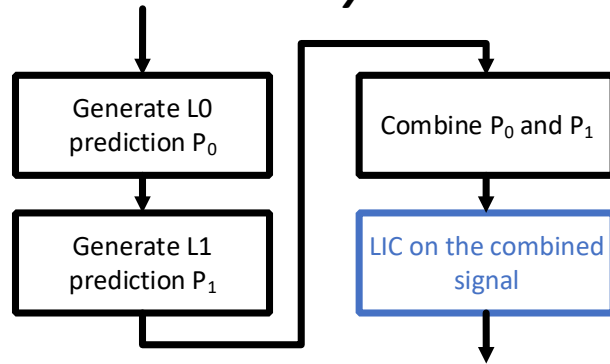
LIC in JEM-7.0

- For bi-prediction, LIC parameters are estimated and applied to L0 and L1 separately
- LIC parameters are estimated at both MC and OBMC



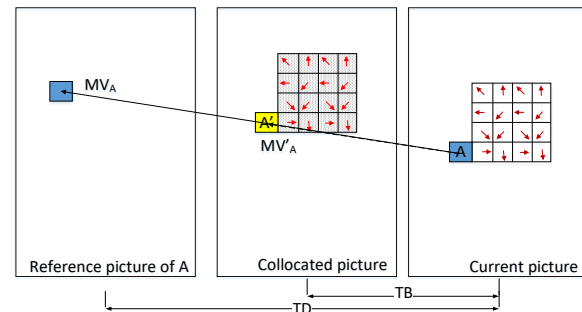
LIC in this response

- LIC parameters are estimated and applied to the bi-prediction signal **only once**
- Reuse the LIC parameters of the MC at OBMC
- ***Reduce the worst-case number of LIC derivations by almost 10 times***



Sub-block merge mode simplification

- ATMVP/STMVP in JEM-7.0
 - Uses fixed 4x4 sub-block size
 - Needs to store 4x4 motion field for all reference pictures
- Simplified ATMVP-based motion derivation
 - Use the same collocated picture concept as in HEVC
 - *Only need to access motion from one reference picture (i.e., the collocated picture) for temporal motion derivation*
 - MV candidates are scaled to the collocated picture in the ATMVP motion field
- Add picture/slice-level signaling to support variable sub-block sizes for ATMVP/STMVP-based motion derivation
 - Encoder determines the optimal sub-block size based on the sizes of sub-block merge CUs in the previous picture at the same temporal layer



Adaptive loop filter simplification

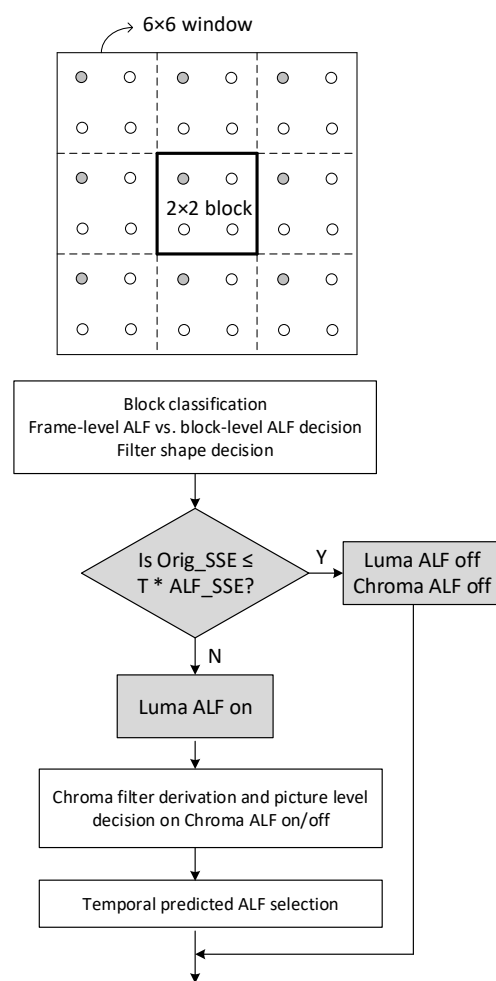
- Simplified block classification
 - For the highest temporal layer pictures, calculate the derivative only at top-left sample position of each 2×2 block
- Encoder-only condition to bias toward disabling ALF for some pictures
 - Luma and chroma ALF are disabled for pictures *in the highest and second highest temporal layers* if:

$$Orig_SSE \leq T \cdot ALF_SSE$$

Orig_SSE: SSE between original picture and reconstructed picture without ALF

ALF_SSE: SSE between the original picture and reconstructed picture with ALF

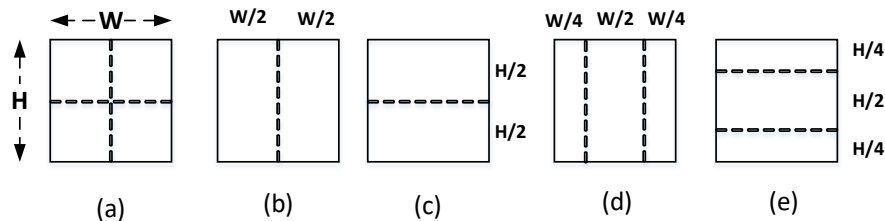
T: bias factor



SDR Category: additional tools not in JEM

Multi-type tree (MTT) partition

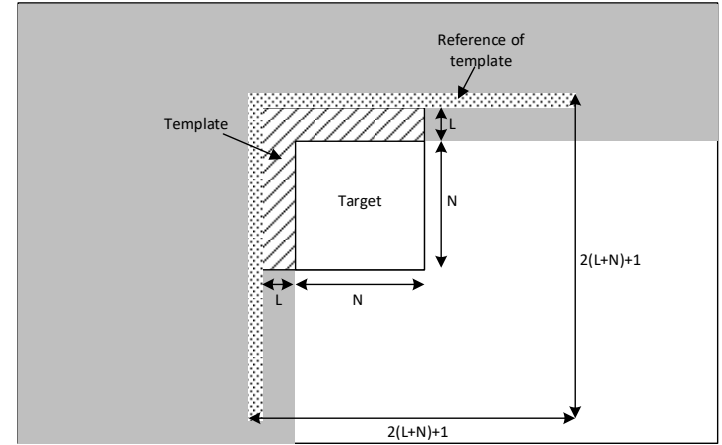
- Besides the quad-tree plus binary-tree (QTBT), triple-tree (TT) from JVET-D0117 is used in this response
 - TT splits one block horizontally/vertically into three parts in a ratio of 1: 2: 1
 - When splitting one QT leaf node, one additional flag is signal to indicate whether it is BT or TT partition
 - Constraints are applied in signaling to avoid redundant partitions



- Fast MTT encoding
 - RD tests of certain repeated partitions are skipped
 - Early termination of horizontal or vertical TT partition based on the RD costs of BT partitions
 - Detailed description available in J0015

Decoder-side intra mode derivation (DIMD)

- DIMD (JVET-C0061 and JVET-D0097) improves intra coding efficiency
 - DIMD flag is signaled at CU level
 - If DIMD, the intra mode is not signaled, and instead derived from neighboring reconstruction samples
 - 129 angular intra directions are implicitly supported
- During parsing, the intra mode of a DIMD CU is unknown, so intra mode propagation (JVET-D0098) is applied to avoid parsing dependency
 - The DIMD intra modes are only derived at reconstruction stage
 - ***Lower parsing complexity than the MPM-based parsing of normal intra mode***



Compression performance

Category	Sequence	HM		JEM	
		CS1	CS2	CS1	CS2
SDR-A	FoodMarket4	35.70%		3.71%	
	CatRobot1	41.50%		3.78%	
	DaylightRoad2	42.87%		4.40%	
	ParkRunning3	34.88%		5.59%	
	Campfire	37.12%		4.44%	
SDR-B	BQTerrace	33.45%	28.90%	4.49%	3.78%
	RitualDance	29.80%	24.32%	3.49%	3.38%
	MarketPlace	30.98%	23.62%	2.65%	3.74%
	BasketballDrive	33.75%	28.26%	3.88%	3.65%
	Cactus	37.09%	30.80%	3.40%	3.65%
Average all		35.72%	27.18%	3.98%	3.64%
Enc Time		1710%	1827%	205%	203%

Decoding complexity: V.S. HM anchor

Category	Sequence	CS1				CS2			
		R1	R2	R3	R4	R1	R2	R3	R4
SDR-A	FoodMarket4	232%	236%	232%	227%				
	CatRobot1	257%	284%	274%	251%				
	DaylightRoad2	231%	241%	251%	257%				
	ParkRunning3	267%	262%	267%	278%				
	Campfire	238%	234%	234%	244%				
SDR-B	BQTerrace	203%	218%	242%	241%	211%	232%	238%	246%
	RitualDance	299%	276%	281%	274%	295%	294%	307%	315%
	MarketPlace	277%	277%	288%	272%	283%	296%	306%	311%
	BasketballDrive	327%	316%	300%	304%	309%	313%	322%	333%
	Cactus	306%	305%	307%	275%	383%	375%	375%	355%
Average		263%				301%			

- Average decoding: 2.63x and 3.01x the HM anchor for CS1 and CS2
- More time saving for CS1 because BIO is more frequently invoked in CS1

Decoding complexity: V.S. JEM anchor

Category	Sequence	CS1				CS2			
		R1	R2	R3	R4	R1	R2	R3	R4
SDR-A	FoodMarket4	27%	28%	29%	32%				
	CatRobot1	27%	28%	28%	29%				
	DaylightRoad2	30%	31%	33%	35%				
	ParkRunning3	33%	32%	33%	35%				
	Campfire	31%	35%	38%	43%				
SDR-B	BQTerrace	26%	28%	32%	33%	39%	45%	46%	49%
	RitualDance	37%	39%	41%	44%	44%	46%	50%	52%
	MarketPlace	31%	33%	35%	36%	41%	41%	44%	46%
	BasketballDrive	36%	38%	41%	44%	44%	47%	49%	54%
	Cactus	30%	31%	32%	32%	39%	41%	45%	48%
Average		33%				45%			

- Average decoding: 0.33x and 0.45x the HM anchor for CS1 and CS2

“Tool-on” and “Tool-off” tests (1)

- Decoding complexity reduction
 - **FRUC SIMP**: simplifications on FRUC
 - **BIO SIMP**: simplifications on BIO
 - **OBMC SIMP**: simplifications on OBMC

Tested tool	“Tool-off”			“Tool-on”		
	BD-rate	Run time		BD-rate	Run time	
		Enc	Dec		Enc	Dec
FRUC SIMP	0.00%	105%	156%	0.14%	95%	67%
BIO SIMP	-0.17%	104%	146%	0.13%	95%	79%
OBMC SIMP	-0.03%	104%	120%	0.05%	93%	87%

These simplifications reduce decoding as well as encoding complexity
@ no performance loss

“Tool-on” and “Tool-off” tests (2)

- Coding performance improvement
 - **MTT**: multiple-type tree
 - **DIMD**: decoder-side intra mode derivation

Tested tool	“Tool-off”			“Tool-on”		
	BD-rate	Run time		BD-rate	Run time	
		Enc	Dec		Enc	Dec
MTT	4.35%	43%	101%	-3.26%	236%	102%
DIMD	0.41%	90%	99%	-0.41%	107%	100%

Further simplifications after submission

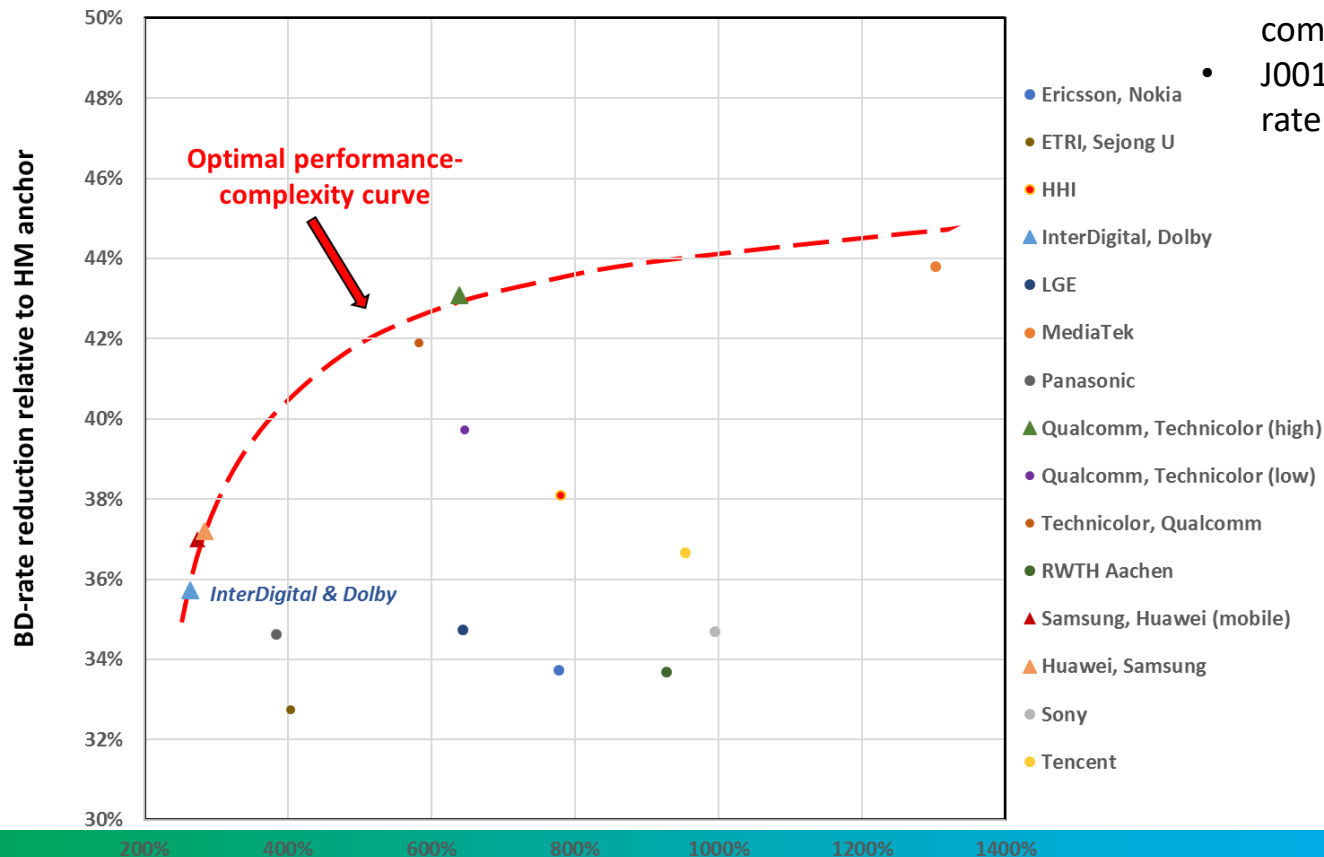
- Further simplification on LIC
 - In the CfP submission, LIC parameter derivation is simplified, but the LIC sample adjustment is still applied at both MC and OBMC stages
 - After CfP submission, LIC is further simplified to apply the LIC sample adjustment only once on the prediction signal after OBMC
 - Significantly reduce both average and worst-case LIC complexity without any performance loss
- Latency reduction on template-based inter prediction (JVET-J0045)
 - Two template-based inter prediction modes in JEM-7.0
 - FRUC template-matching and LIC
 - Significant decoding architecture issue because the reconstructed samples of spatial neighbors are needed in these template-based inter modes
 - Inhibits the ability to parallel decode CTUs or CUs
 - Proposed solution in J0045:
 - Use prediction samples and add reconstructed DC component only to form the template samples
 - Keep majority of the coding gain of these two modes: only lose 0.46% of the total 3.27% BD rate savings from these two tools

SDR summary

- Design goals of this CfP response
 - Reduce average and worst-case decoding complexity without causing performance loss
 - Further improve the coding efficiency beyond JEM
- The response provides good candidate technologies to achieve the goals
 - Average coding efficiency about 35.7% and 4% better than HM and JEM anchors
 - Average decoding time about 2.6x and 0.3x the decoding time of HM and JEM
 - Tool-on and tool-off tests show significant complexity reduction and negligible performance loss from the proposed simplifications
- The proposed technologies are suitable to be included in the Test Model
 - Based on well-known JEM tools, commonalities with other proposals
 - Good complexity-performance trade-off

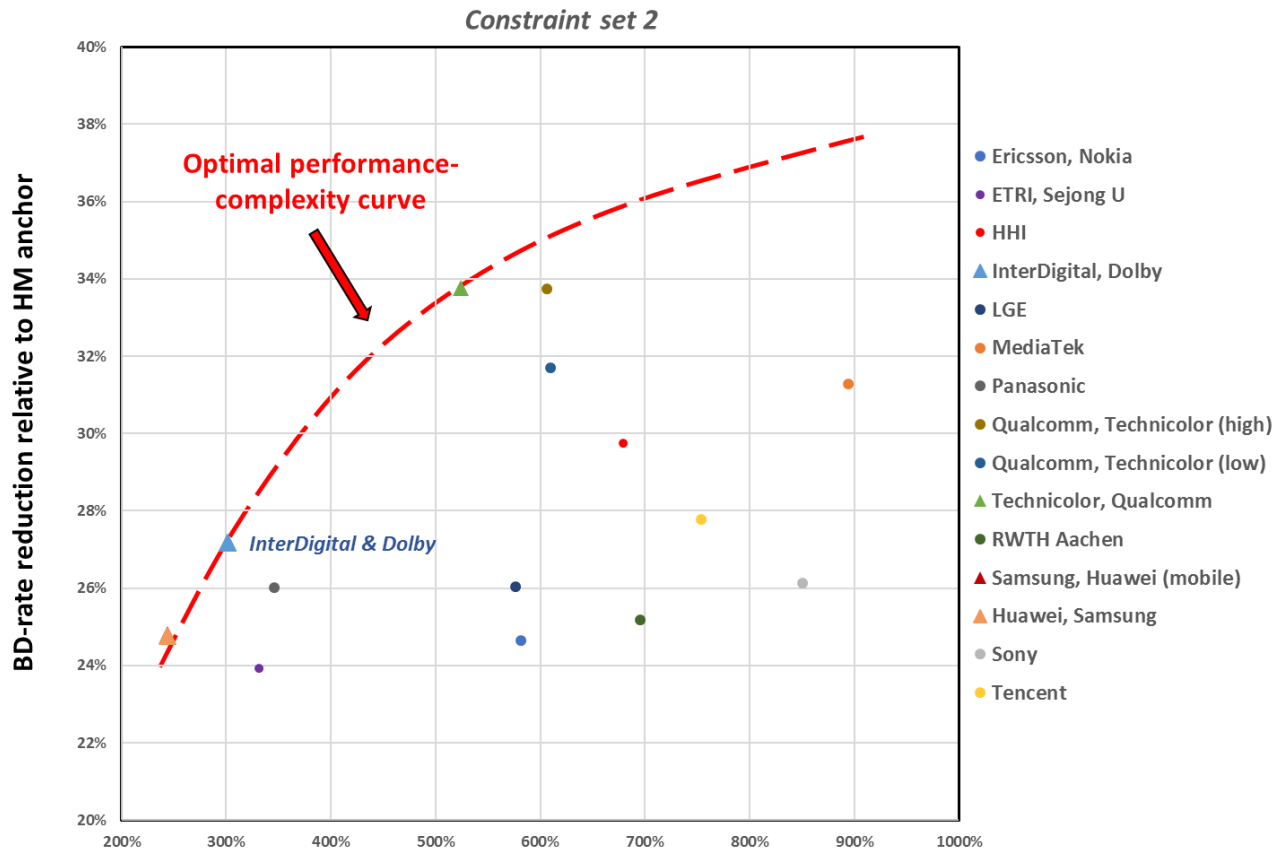
Analysis of 15 SDR responses : CS1

Constraint set 1



- 14 responses with decoding complexity < 10x of HM
- J0018 which has the best BD-rate performance

Analysis of 15 SDR responses : CS2



HDR Summary

- Two major HDR specific coding tools
 - In-loop reshaping
 - Luma-based QP prediction
- Pre- and Post-processing are unnecessary
 - All proposed tools reside inside core codec
 - Keep conformance point same as SDR
- HDR CfP submitted bitstreams target better subjective performance
 - In-loop reshapers can also be configured to maximize HDR objective metrics
- HDR tools should be part of the base codec:
 - Wide commercial success of HDR in the market place.

360° summary

- Projection format to improve coding efficiency
 - Hybrid Angular Cubemap (HAC)
 - Adaptive Frame Packing (AFP)
- Coding tools
 - Geometry Padding of Reference Pictures (GP)
 - Face Discontinuity Handling (FDH)
 - Relatively simple modifications to SDR coding tools can resolve many seam artifacts problems
 - Padding, guard bands, blending, etc. are not sufficient
- Post-processing
 - Post-Filtering
- 360 tools and of SDR tools should be part of the future video coding standards

THANK YOU