KWANGWOON UNIVERSITY

Scalable structures and inter-layer predictions for HEVC scalable extension

JCTVC-F096

H.M. Choi, J. Nam, D. Sim

Image Processing Systems Laboratory

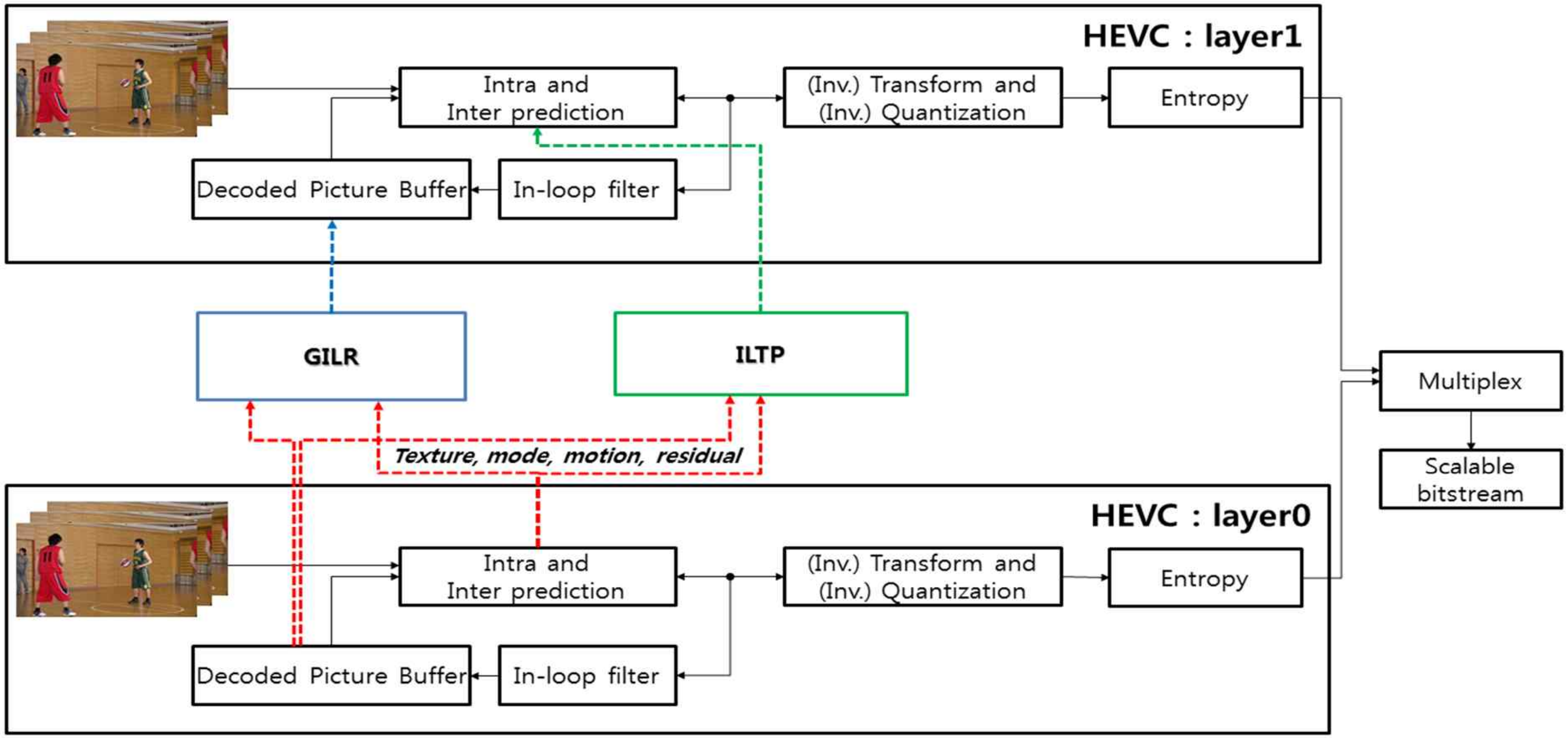
Status:	Input Document to JCT-VC
Purpose:	Investigation

Introduction

- HEVC scalable coding
  - HEVC scalable coding would be a valuable standard to support various services in the mobile, UHD and 3D markets
- Proposed scalable structures with HM2.0
  - Simulcast approach with two different loop designs
    - Single-loop
    - Multi-loop
  - New two inter-layer predictions
    - Inter-layer texture prediction (ILTP)
    - Generalized inter-layer reference frame (GILR)

Simulcast scalable video coder with HM2.0 JCTVC-F096

- Proposed scalable system
  - Each layer employs HEVC tools
  - Support spatial and temporal scalabilities
  - Inter-layer predictions

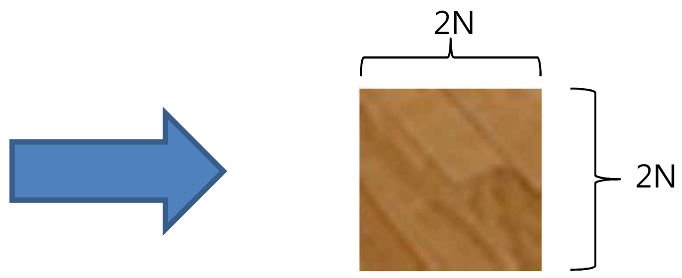


Inter-layer texture prediction (ILTP)

- ILTP is a scheme to predict a PU at an enhancement layer from corresponding block of the reference layer
  - ILTP makes a texture signal for 2Nx2N PU size
  - The way to make the texture is different depending on loop designs
    - Three ways in predicting the PU depending on slice types and coding modes for the single-loop design
    - One way to predict the PU for the multi-loop design as the same way of the single-loop for intra-slices.

Single-loop design
* intra slice at Enh.
* inter slice at Enh.
  -intra block at Ref.
  -inter block at Ref.

Multi-loop design



2Nx2N predictor

ILTP for the single-loop design

- For intra slices in the enhancement layer
    - Reconstruct a corresponding block of the reference layer
    - Perform up-sampling

	Filter coefficients
Luma	{–1, 4, –11, 40, 40, –11, 4, –1}
Chroma	{–4, 36, 36, –4}

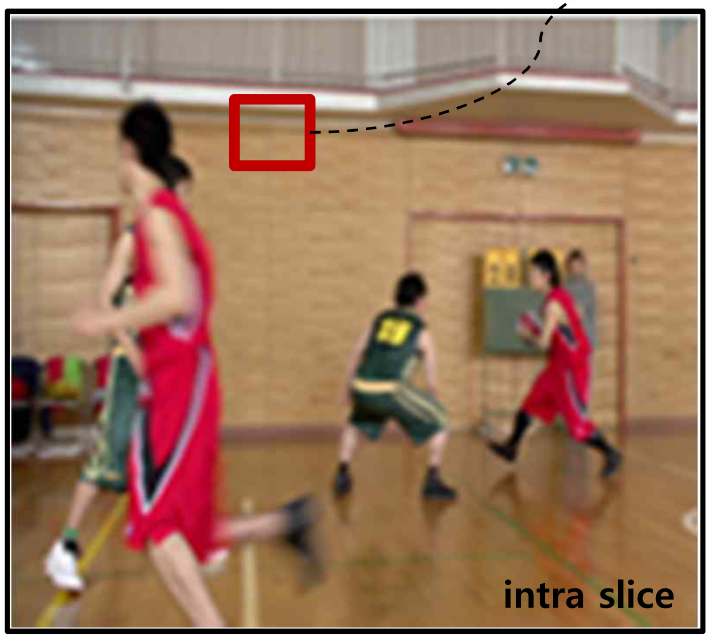
- For Inter slice in the enhancement layer
  - Intra block in the reference layer
    - Parse intra mode of the corresponding block in the reference layer
    - Perform an intra-prediction with <u>the decoded intra mode and the neighboring decoded pixels in the enhancement layer</u>
    - Update an up-sampled residual signal of the reference layer
  - Inter block in the reference layer
    - Decode MV, reference index of the corresponding block in the reference layer
    - Motion compensation at the enhancement layer
    - Update an up-sampled residual signal of the reference layer

Intra slices in the enhancement layer for single-loop

➔ After coding the base layer

3.Fetch for 2Nx2N predictor

- **Send one flag for indicating an ILTP**



enhancement layer

2.Up-sampling

base layer

1.Reconstruction

For half-pel position

	Filter coefficients
Luma	{–1, 4, –11, 40, 40, –11, 4, –1}
Chroma	{–4, 36, 36, –4}

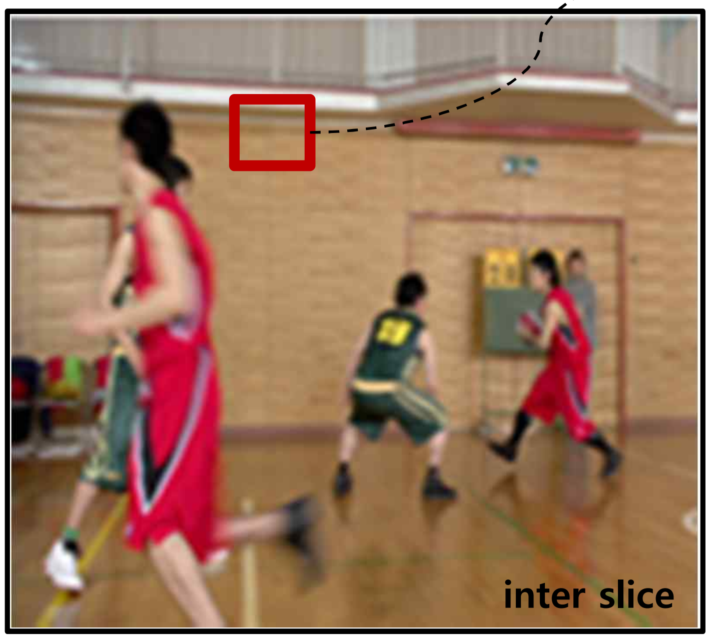
🡐 Before coding the enhancement layer

Inter slices in the enhancement layer for single-loop

➔ After coding the base layer

3.Fetch for 2Nx2N predictor

- **Send one flag for indicating an ILTP**



enhancement layer

2. Add a resi.

1. Fill up a frame

inter slice

base layer

Corresponding Block (C.B) == Intra block

- Parse an intra mode of C.B
- Perform an intra prediction with the parsed mode in enh. layer (Using neighboring pixels at enh. layer)

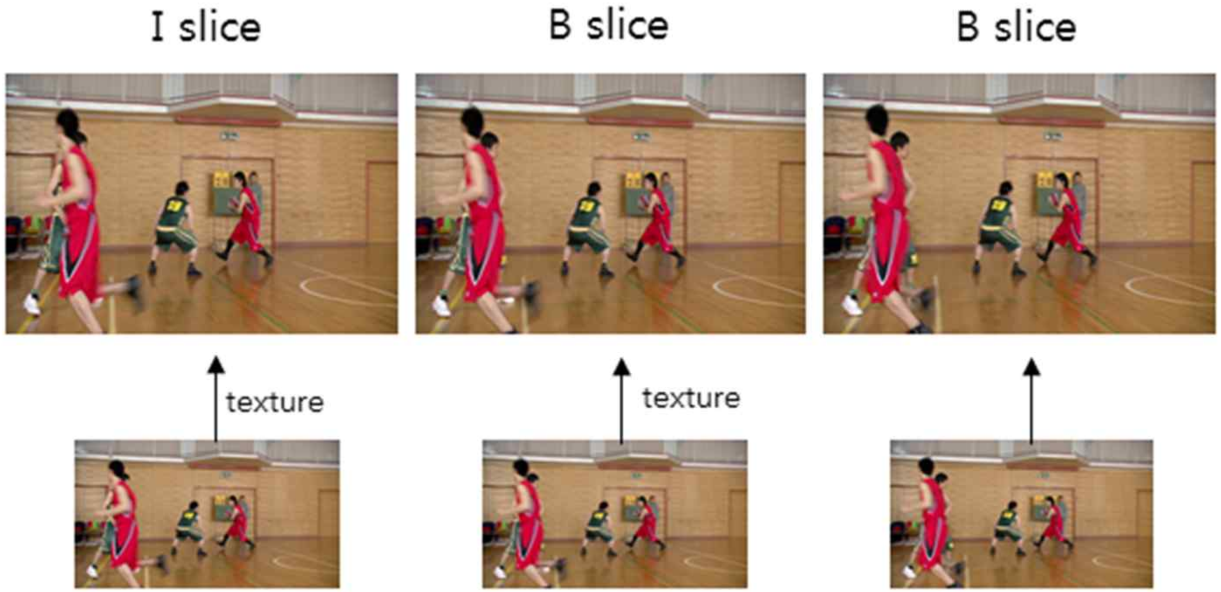
Corresponding Block (C.B) == Inter block

- Decode MV, reference index of C.B
- Scale-up MV
- Motion Compensation at enh. layer

🡐 Before coding the enhancement layer

ILTP for the multi-loop design

- For multi-loop case
  - The reference layer is to be fully reconstructed.
  - ILTP works in the same way of the single-loop for intra-slices



ILTP for the multi-loop design

➔ After coding the base layer

3.Fetch for 2Nx2N predictor

- **Send one flag for indicating an ILTP**

enhancement layer

2.Up-sampling

base layer

1.Reconstruction

For half-pel position

	Filter coefficients
Luma	{–1, 4, –11, 40, 40, –11, 4, –1}
Chroma	{–4, 36, 36, –4}

🡐 Before coding the enhancement layer

Generalized inter-layer reference frame (GILR) JCTVC-F096

- GILR generated from the reference layer is a reference frame to be inserted into DPB of the enhancement layer
  - Intra slices of the enhancement layer are coded by inter prediction with referring to GILR as a reference frame.
    - The Intra slice type are changed into inter slice type
    - Such as slices are called G-I slice
  - For <u>inter slices of the enhancement layer, </u>GILR is added into DPB list 0 and list 1.
    - Number of reference frame does not increase due to the GILR.
  - Each block of the GILR is generated with the same way of the ILTP.

Single-loop design

Multi-loop design

Generalized reference frame

GILR for the single-loop design

- GILR for intra slices in the enhancement layer
    - Reconstruct a corresponding block of the reference layer
    - Interpolation (DCT-IF 8-tap for Luma, DCT-IF 4-tap for Chroma)
    - Insert the GILR into DPB of the enhancement layer

- GILR for Inter slices in the enhancement layer
  - For Intra blocks in the reference layer
    - Parse intra mode of the corresponding block in the reference layer
    - Perform an intra-prediction with <u>the decoded intra mode and the neighboring decoded pixels in the enhancement layer</u>
    - Update an up-sampled residual signal of the reference layer
  - For Inter blocks in the reference layer
    - Decode MV, reference index of the corresponding block in the reference layer
    - Motion compensation at enhancement layer
    - Update an up-sampled residual signal of reference layer

  Insert GILR into DPB of enhancement layer

Intra slices in the enhancement layer for single-loop

→ After coding the base layer

3. Insert into DPB of enhancement layer

- **No differences in terms of syntax, compared with HM2.0**

enhancement layer

2.Up-sampling

base layer

1.Reconstruction

For half-pel position

	Filter coefficients
Luma	{–1, 4, –11, 40, 40, –11, 4, –1}
Chroma	{–4, 36, 36, –4}

← Before coding the enhancement layer

Inter slices in the enhancement layer for single-loop

➔ After coding the base layer

3. **Insert into DPB of enhancement layer**

- **No differences in terms of syntax, compared with HM2.0**

enhancement layer

2. Add a resi.

1. Fill up a frame

inter slice

base layer

Corresponding Block (C.B) == Intra block

- Parse an intra mode of C.B
- Perform an intra prediction with the parsed mode in enh. layer (Using neighboring pixels at enh. layer)

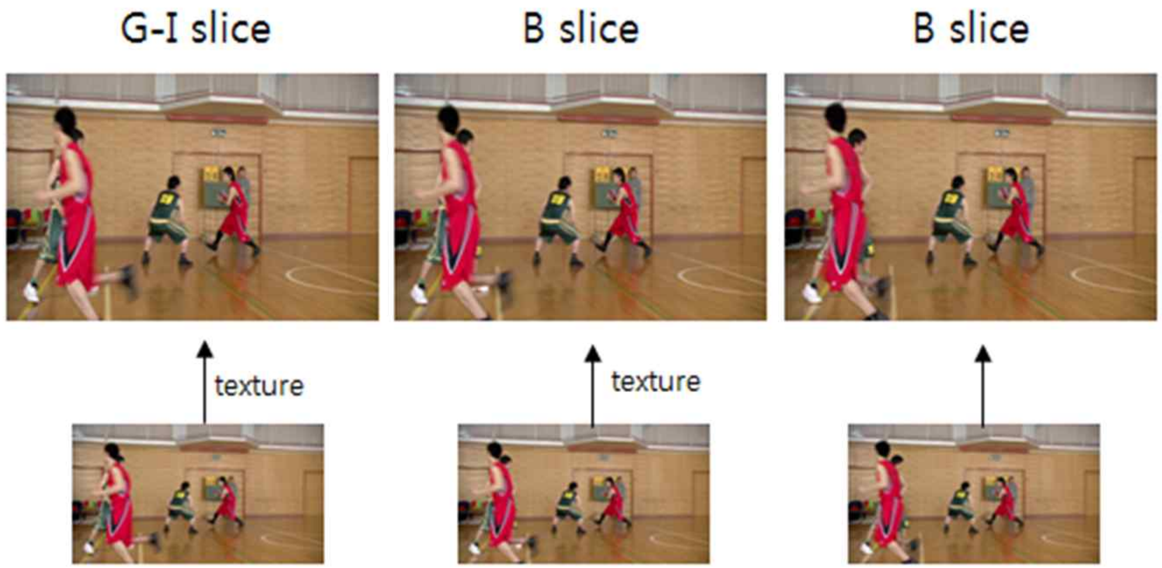
Corresponding Block (C.B) == Inter block

- Decode MV, reference index of C.B
- Scale-up MV
- Motion Compensation at enh. layer

← Before coding the enhancement layer

GILR for the multi-loop design

- For multi-loop case
  - The reference layer is to be fully reconstructed.
  - GILR works in the same way of the single-loop for intra-slices



GILR for the multi-loop design

➔ After coding the base layer

3.Insert into DPB of enhancement layer

- **No differences in terms of syntax, compared with HM2.0**

enhancement layer

2.Up-sampling

base layer

1.Reconstruction

For half-pel position

	Filter coefficients
Luma	{–1, 4, –11, 40, 40, –11, 4, –1}
Chroma	{–4, 36, 36, –4}

← Before coding the enhancement layer

Test conditions

- Anchor : HM2.0
- Proposed scalable codec: implemented based on HM2.0

- Tested configurations
  - Single-loop scalable structure with ILTP
  - Single-loop scalable structure with GILR
  - Multi-loop scalable structure with ILTP
  - Multi-loop scalable structure with GILR

- Common test condition
  - All intra, low delay, random access in High Efficiency (HE)

- Common test sequence
  - Class B, Class C, Class D

Test condition (cont'd)

- Suppose two layers in spatial scalability
  - Base layer : generated by the JSVM "DownConverter" with the dyadic factor
  - Enhancement layer : use the common test sequences such as Class B, C and D

- QP (22, 27, 32, 37) are applied for two layers with same value

Evaluation measures

- Coding gain evaluation
  - Compare BD-bitrate reduction only for the enhancement layer of simulcast and the proposed scalable system
  - According to the document "Recommended Requirements and Test Points for Scalable Coding Extension"

$$\textit{Coding Gain (CG)} = 1/(1 - BR/(\alpha * R)) = \alpha / \gamma$$

	Simulcast	Scalable system
Layer 0	R	β*R
Layer 1	α*R	γ*R

* the difference between R and β*R is negligible in this contribution

- Coding performance evaluation
  - compare the enhancement layer of simulcast with the total layer of scalable system in terms of BD bitrate

$$CP = \alpha / (\beta + \gamma)$$

Performance of all intra in HE

- CG evaluation

Anchor : HM2.0

	Single with ILTP			Single with GILR			Multi with ILTP			Multi with GILR		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Class B	-18.6	-12.8	-10.0	-29.9	-20.0	-16.8	-18.6	-12.8	-10.0	-29.9	-20.0	-16.8
Class C	-10.7	-6.6	-4.9	-17.4	-7.4	-5.1	-10.7	-6.6	-4.9	-17.4	-7.4	-5.1
Class D	-12.8	-8.2	-7.8	-15.3	-2.9	-1.4	-12.8	-8.2	-7.8	-15.3	-2.9	-1.4
Avg	**-14.04**	**-9.22**	**-7.56**	**-20.86**	**-10.09**	**-7.77**	**-14.04**	**-9.22**	**-7.56**	**-20.86**	**-10.09**	**-7.77**

- CP evaluation

	Single with ILTP			Single with GILR			Multi with ILTP			Multi with GILR		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Class B	17.6	26.0	29.8	13.4	28.3	32.6	17.6	26.0	29.8	13.4	28.3	32.6
Class C	21.3	26.7	29.0	19.7	33.4	36.7	21.3	26.7	29.0	19.7	33.4	36.7
Class D	17.4	23.3	24.0	20.4	36.8	38.9	17.4	23.3	24.0	20.4	36.8	38.9
Avg	**18.78**	**25.31**	**27.58**	**17.84**	**32.86**	**36.06**	**18.78**	**25.31**	**27.58**	**17.84**	**32.86**	**36.06**

* ILTP and GILR procedure of all intra case is same for single and multi-loop

Performance of low delay in HE

- CG evaluation

Anchor : HM2.0

	Single with ILTP			Single with GILR			Multi with ILTP			Multi with GILR		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Class B	-5.1	-3.6	-2.5	-4.8	-4.6	-3.2	-6.4	-1.2	0.1	-6.8	-1.2	1.2
Class C	-3.6	-2.5	-2.5	-2.4	-1.6	-1.2	-5.3	-1.0	-0.1	-4.8	0.3	1.6
Class D	-2.4	-1.0	-1.2	-0.2	1.9	2.3	-3.7	0.4	0.3	-1.9	4.8	5.0
Avg	**-3.70**	**-2.36**	**-2.07**	**-2.48**	**-1.43**	**-0.67**	**-5.11**	**-0.60**	**0.07**	**-4.50**	**1.31**	**2.59**

- CP evaluation

	Single with ILTP			Single with GILR			Multi with ILTP			Multi with GILR		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Class B	27.2	29.4	30.8	28.1	28.7	30.5	26.1	33.2	35.0	26.3	34.1	37.2
Class C	27.1	28.5	28.4	32.8	33.5	35.1	25.4	30.9	32.1	26.2	32.9	34.6
Class D	25.3	27.0	26.9	27.8	30.4	30.9	24.1	29.5	29.2	26.2	34.8	34.9
Avg	**26.54**	**28.31**	**28.71**	**29.59**	**30.88**	**32.16**	**25.20**	**31.23**	**32.11**	**26.23**	**33.95**	**35.56**

Performance of random access in HE

- CG evaluation

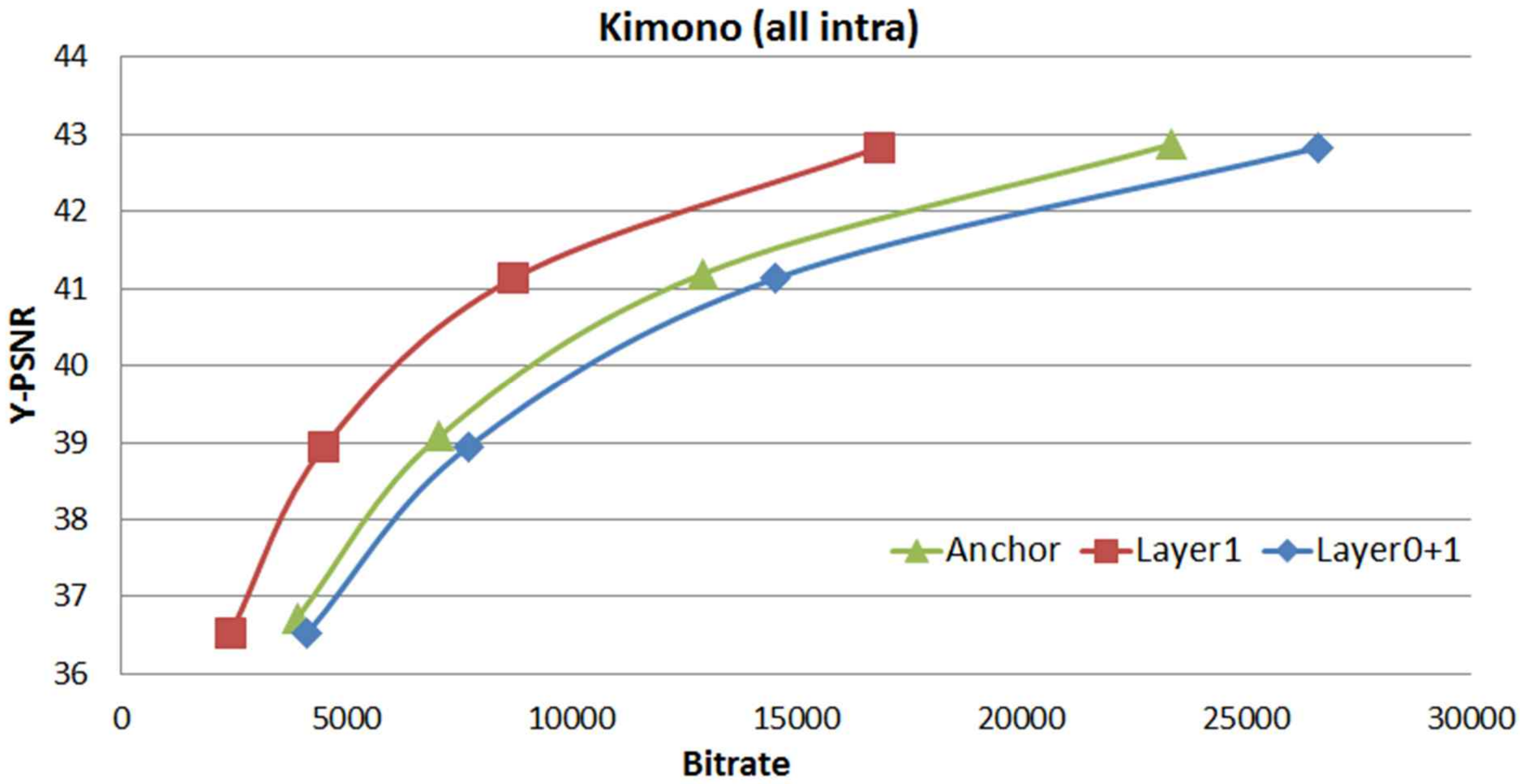
Anchor : HM2.0

	Single with ILTP			Single with GILR			Multi with ILTP			Multi with GILR		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Class B	-14.3	-15.0	-13.7	-13.7	-11.5	-9.2	-16.7	-12.2	-10.5	-16.1	-5.9	-2.3
Class C	-8.8	-8.5	-8.2	-6.2	-3.4	-2.7	-11.3	-66	-5.4	-8.8	0.7	3.1
Class D	-6.4	-5.9	-5.9	-3.3	0.6	1.0	-8.4	-3.8	-4.1	-6.2	4.5	4.5
Avg	**-9.82**	**-9.78**	**-9.27**	**-7.73**	**-4.75**	**-3.65**	**-12.13**	**-7.53**	**-6.66**	**-10.37**	**-0.26**	**1.77**

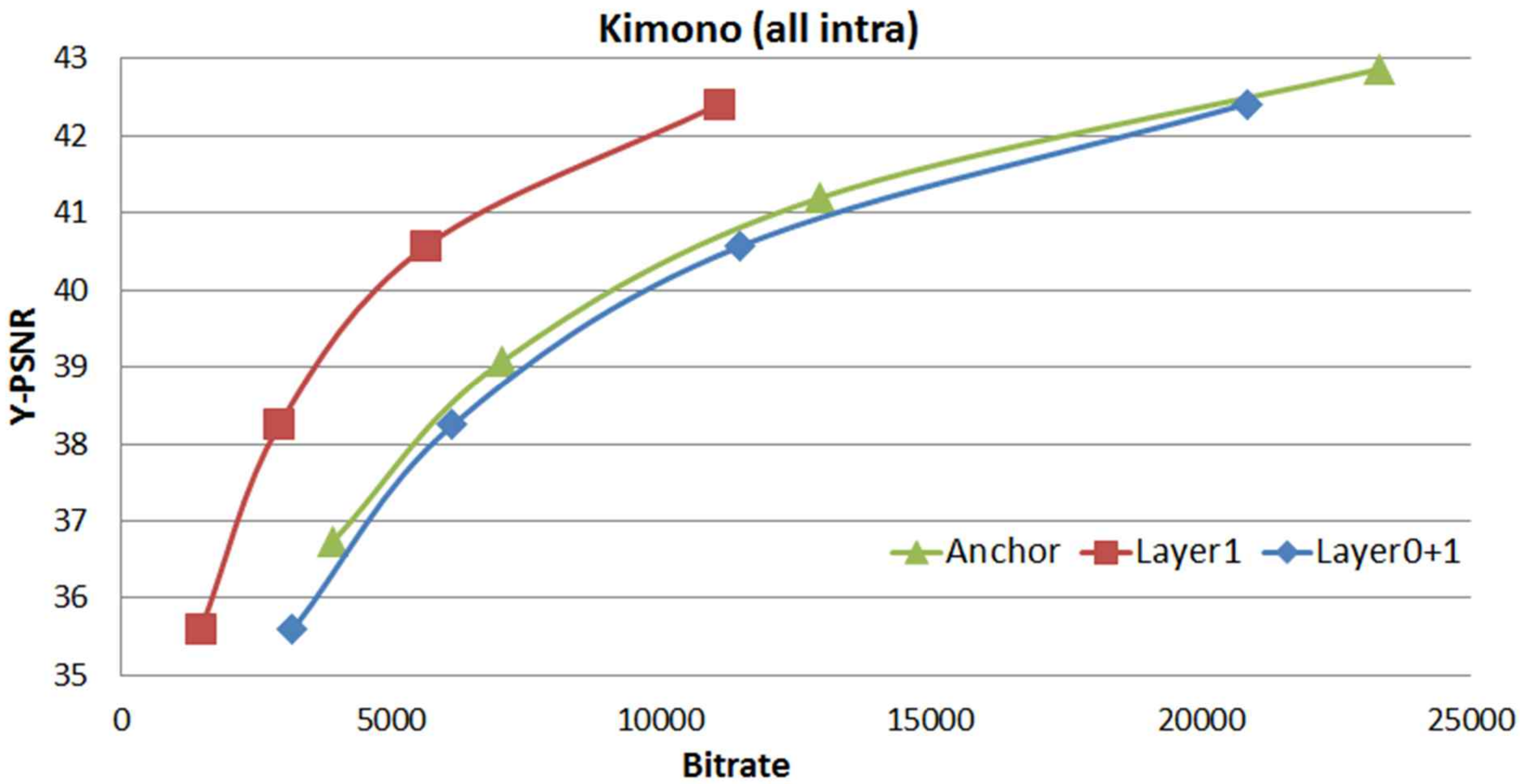
- CP evaluation

	Single with ILTP			Single with GILR			Multi with ILTP			Multi with GILR		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Class B	18.1	17.8	19.3	20.6	24.1	26.7	15.7	22.8	24.8	18.5	33.4	38.2
Class C	21.9	22.3	22.6	26.1	29.6	30.4	19.3	25.5	27.0	23.5	36.1	39.2
Class D	22.3	22.8	22.9	26.7	31.3	31.8	20.4	26.3	26.1	23.8	37.5	37.5
Avg	**20.76**	**20.96**	**21.59**	**24.43**	**28.33**	**29.65**	**18.47**	**24.88**	**25.96**	**21.91**	**35.65**	**38.28**

RD-Curve of Kimono sequence in ILTP



RD-Curve of Kimono sequence in GILR



Conclusion

- Scalable structures and inter-layer predictions were proposed
  - Scalable structures with single and multi-loop design
  - Inter-layer predictions (ILTP/ GILR)
- Coding efficiency in High Efficiency condition (HE)
  - Single-loop with ILTP
    - -14.04% (AI), -3.70% (LD), -9.82% (RA)
  - Single-loop with GILR
    - -20.86% (AI), -2.48% (LD), -7.73% (RA)
  - Multi-loop with ILTP
    - -14.04% (AI), -5.11% (LD), -12.13% (RA)
  - Multi-loop with GILR
    - -20.86% (AI), -4.50% (LD), -10.37% (RA)
- Single and multi-loop design integrated into HEVC scalable extension assist to support spatial scalability with multi-view functionality
- The proposed GILR is quite efficient
  - Better coding efficiency
  - No syntax change, compared with HM2.0
  - GILR is considered as a consolidated solution for multi-view scalability