JCTVC-K0034
SCALABLE CODING TECHNOLOGY PROPOSAL BY INTERDIGITAL COMMUNICATIONS

InterDigital Communications, LLC

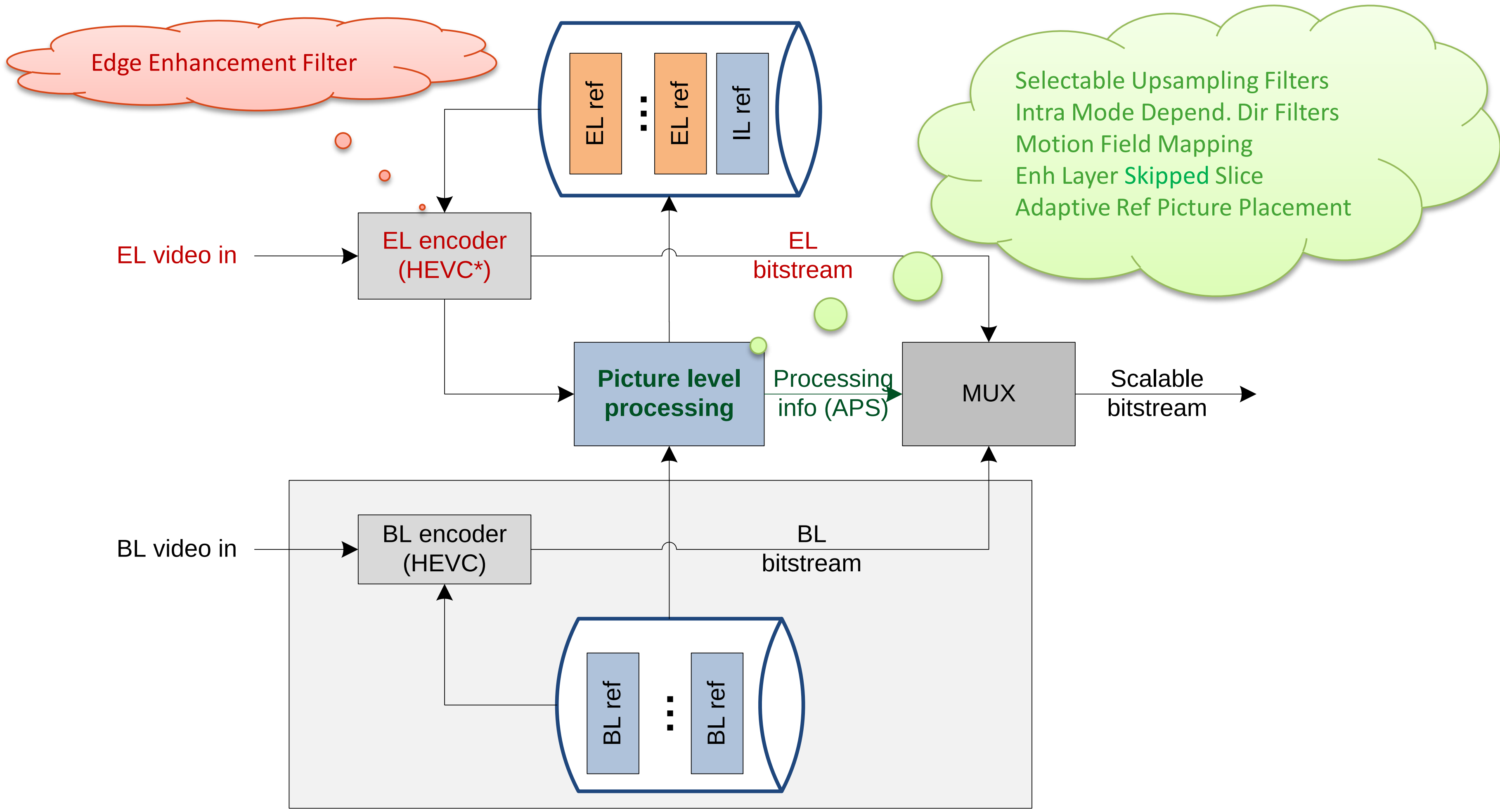
11th JCT-VC meeting, Oct 2012

Introduction

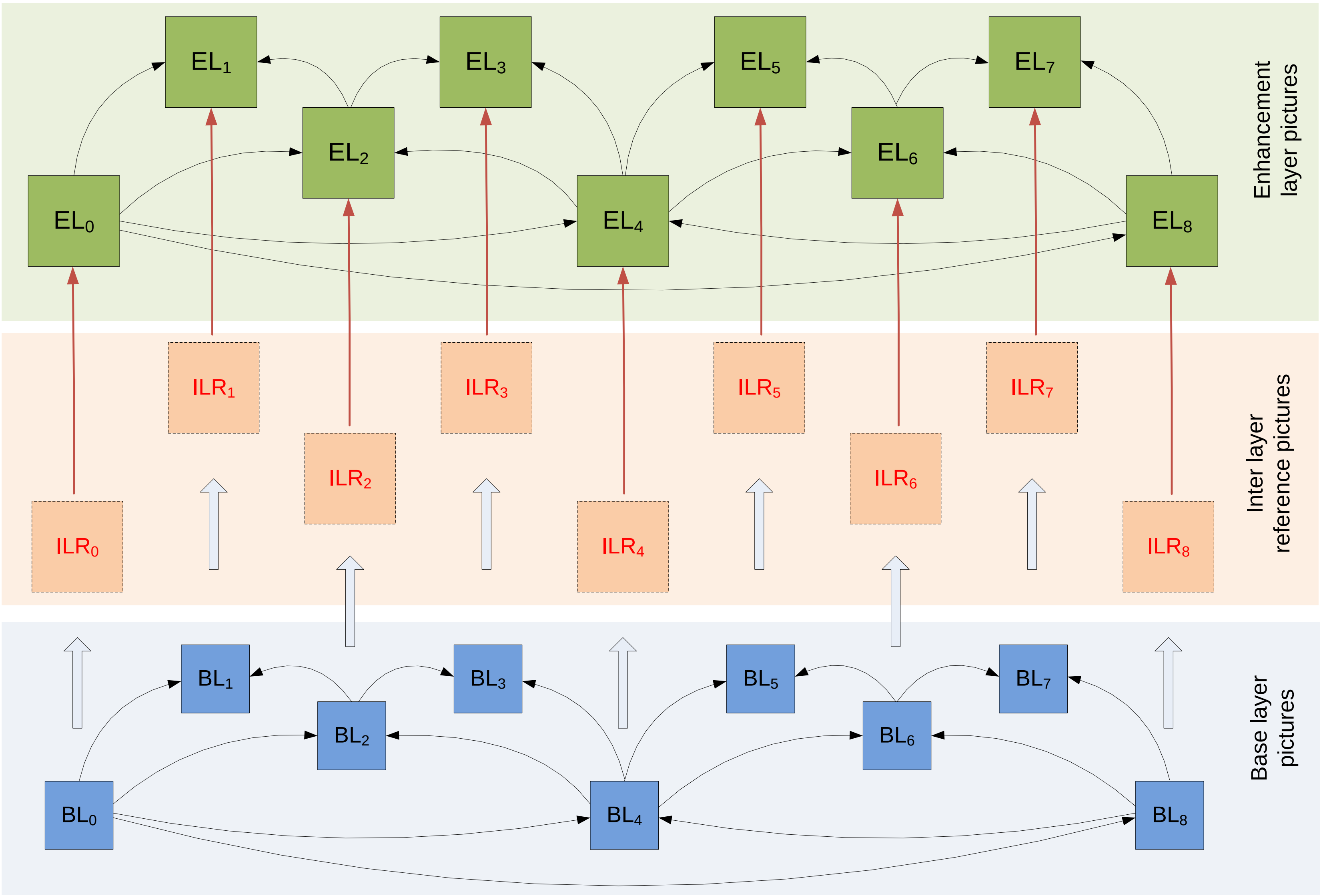
- Our scalable architecture and design philosophy:
  - MVC-like: inter layer reference is added as an additional reference picture for enhancement layer coding
  - Well aligned with 3D extensions in JCT-3V
  - Minimal change to enhancement layer codec
- Primary goal:
  - Improve scalable coding efficiency by maximally re-using current single layer HEVC design
- Our performance data is based on:
  - No change to EL block level operations
  - No change to EL prediction structure
  - No ALF or IBDI
  - No change to EL lambda factor
- Proposed inter layer processing technologies:
  - **SUF:** Selectable Upsampling Filters
  - **IMDDF:** Intra Mode Dependent Directional Filters
  - **MFM:** Motion Field Mapping
  - **ELSkip:** EL skipped slice
  - **ARPP:** Adaptive Reference Picture Placement
- Additional in-loop filter in EL:
  - **EEF:** Edge Enhancement Filter

INTERDIGITAL.

The Proposed Scalable Encoder Block Diagram



Prediction Structure of the Proposed System



INTERDIGITAL.

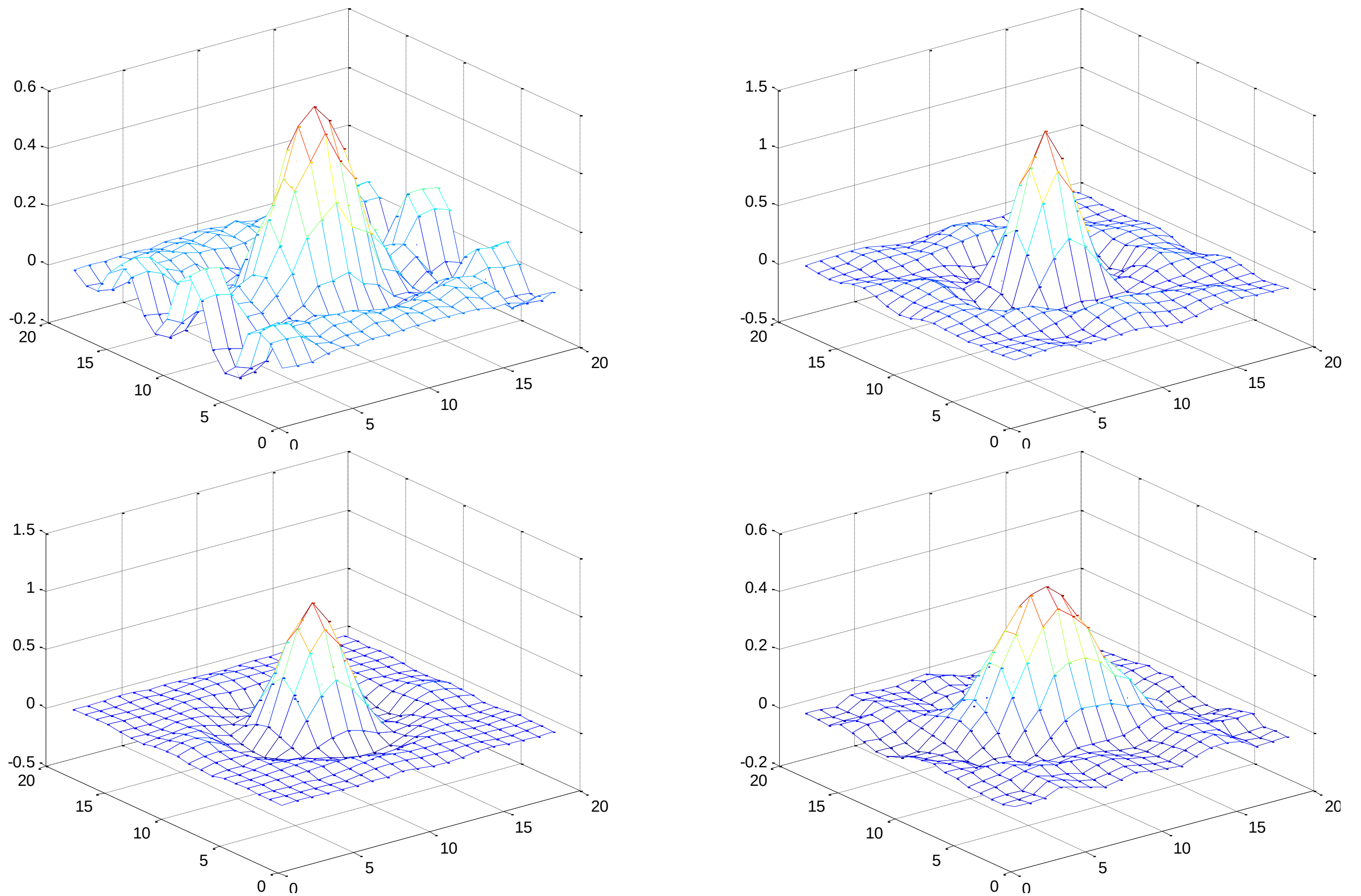
The Default Upsampling Filters

- Gaussian windowed Sinc function
- 2-D separable
- 8-tap luma filters, 4-tap chroma filters
- 8-bit filter coefficient precision
- 16-bit implementation

Luma	2X		[-3	8	-21	80	80	-21	8	-3]/128
	1.5X		[-5	11	-24	107	53	-18	8	-4]/128
			[-4	8	-18	53	107	-24	11	-5]/128
Chroma	2X	Vertical	[-17	100	56	-11]/128				
			[-3	15	126	-10]/128				
		Horizontal	[-15	79	79	-15]/128				
	1.5X	Vertical	[-20	122	38	-12]/128				
			[-20	71	101	-24]/128				
			[-4	12	128	-8]/128				
		Horizontal	[-23	113	54	-16]/128				
			[-16	54	113	-23]/128				

Selectable Upsampling Filters

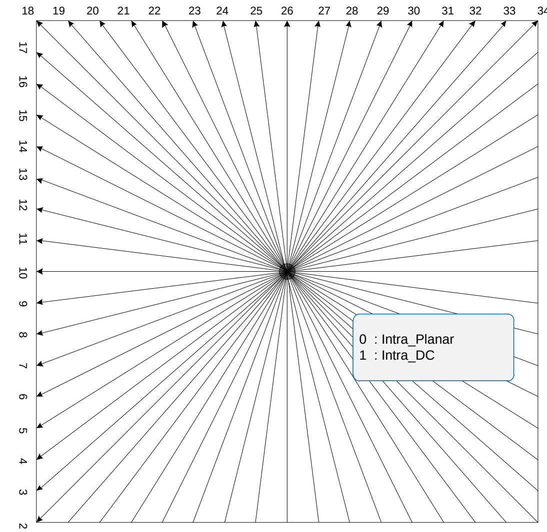
- 4 additional upsampling filters with wider range of filter characteristics
  - Non-separable, with sizes 7x7, 7x6, 6x7, and 6x6, 16-bit implementation
  - Selected on picture level, applied to luma only

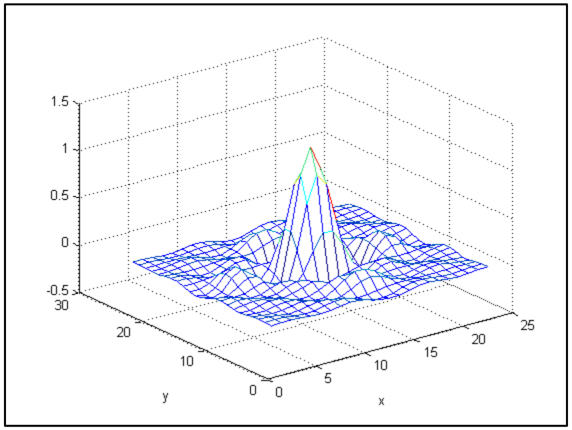


INTERDIGITAL.

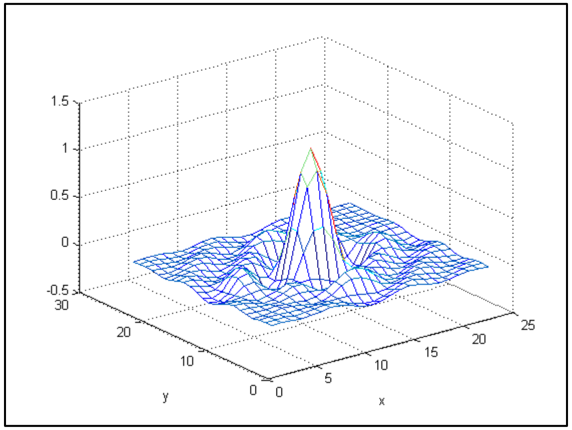
Intra Mode Dependent Directional Filters

- HEVC supports 35 intra prediction modes
- Base layer block intra prediction modes provide directionality information
- IMDDF is a set of poly-phase directional filters for different prediction modes
  - 7x7, 7x8, 8x7 and 8x8
- Applied to luma component only
- Current implementations applies IMDDF on intra coded pictures only



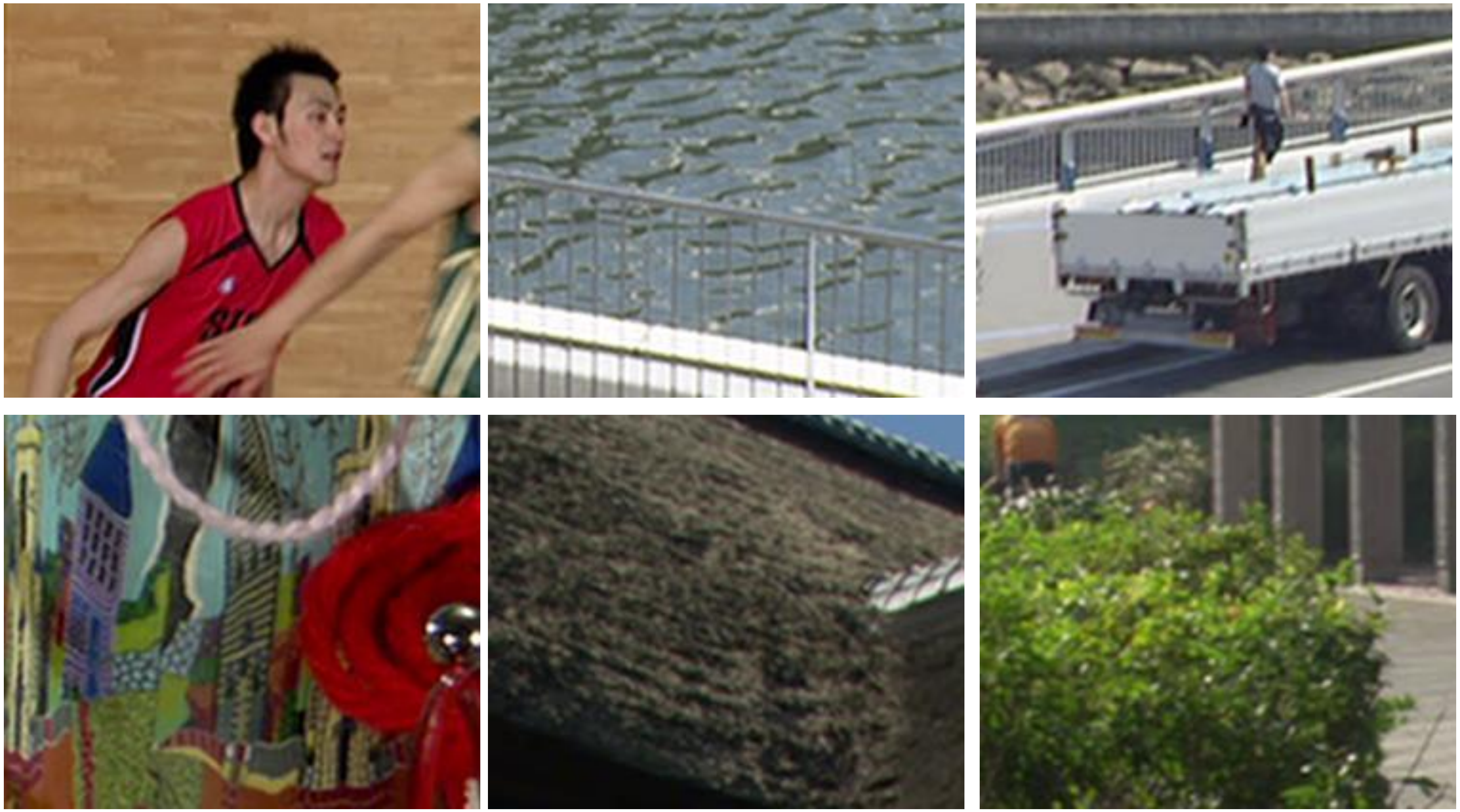


1.5x IMDDF for 135^o diagonal (mode 2)



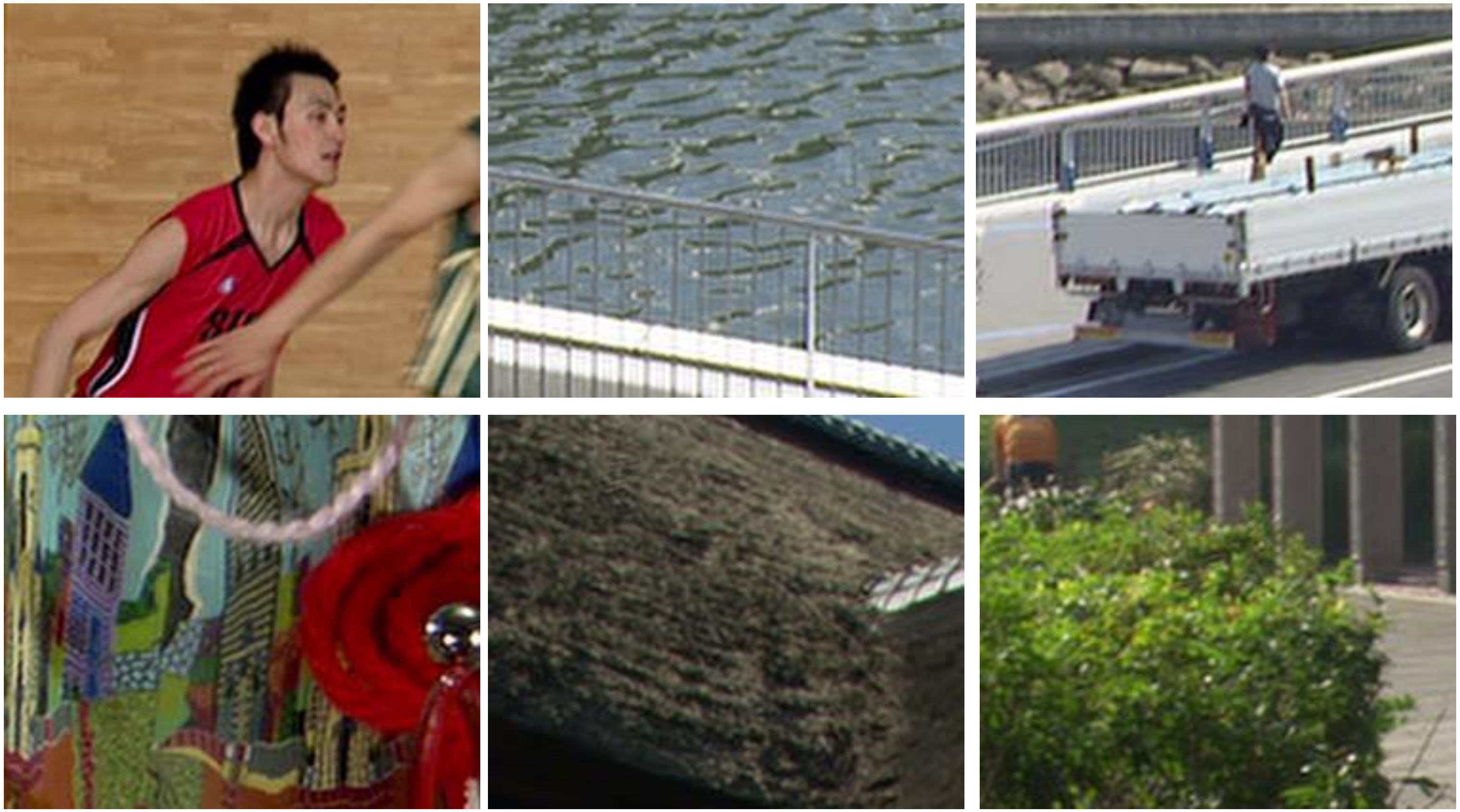
1.5x IMDDF for 45^o diagonal (mode 18)

IMDDF visual examples: fixed 2x



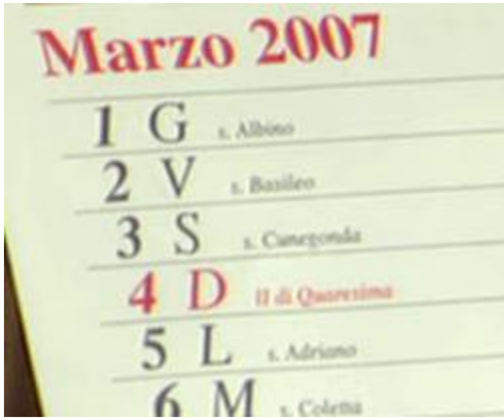
INTERDIGITAL.

IMDDF visual examples: IMDDF 2x



INTERDIGITAL.

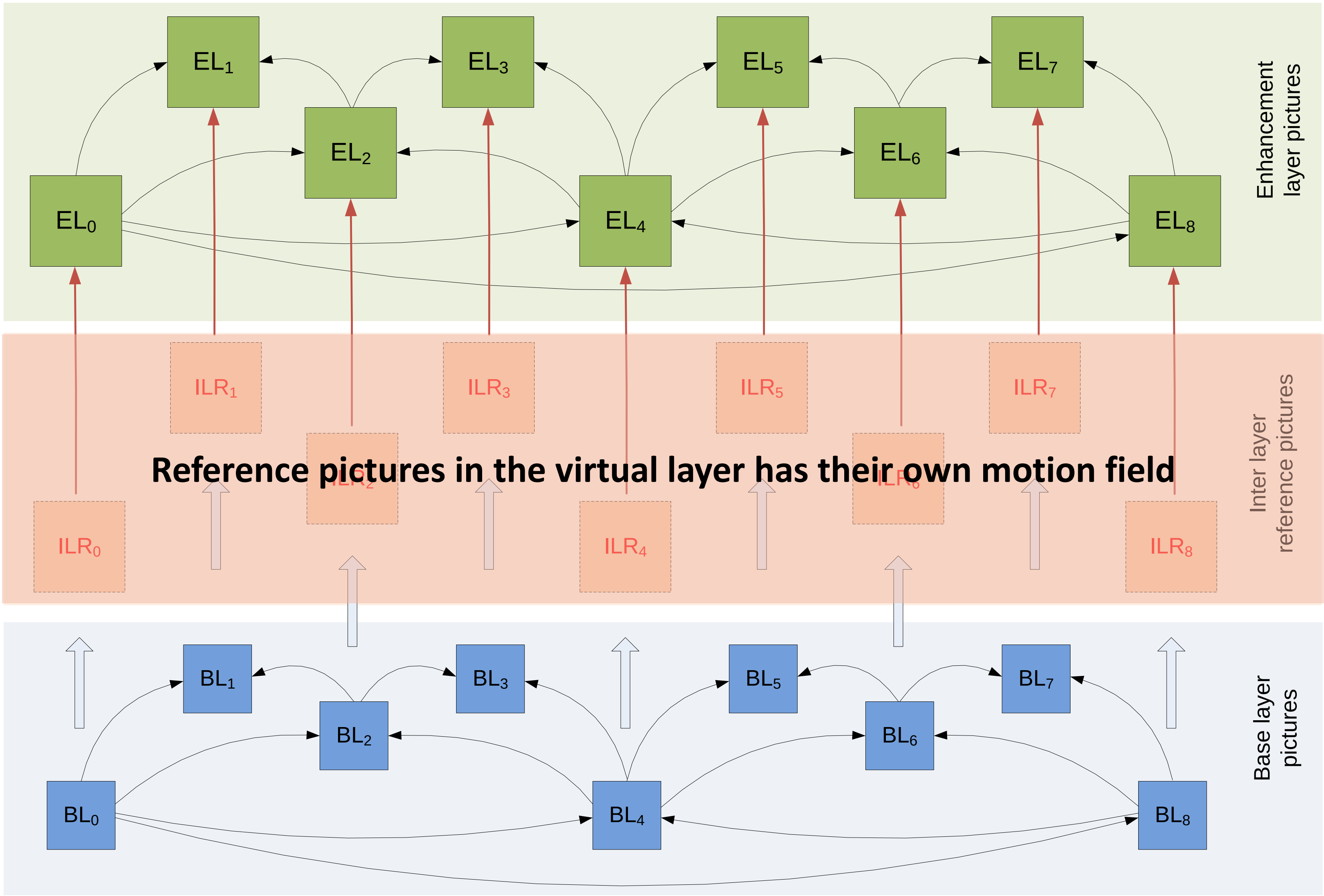
IMDDF visual examples : fixed 1.5x



INTERDIGITAL.

IMDDF visual examples: IMDDF 1.5x

Inter Layer References Form a Virtual Layer



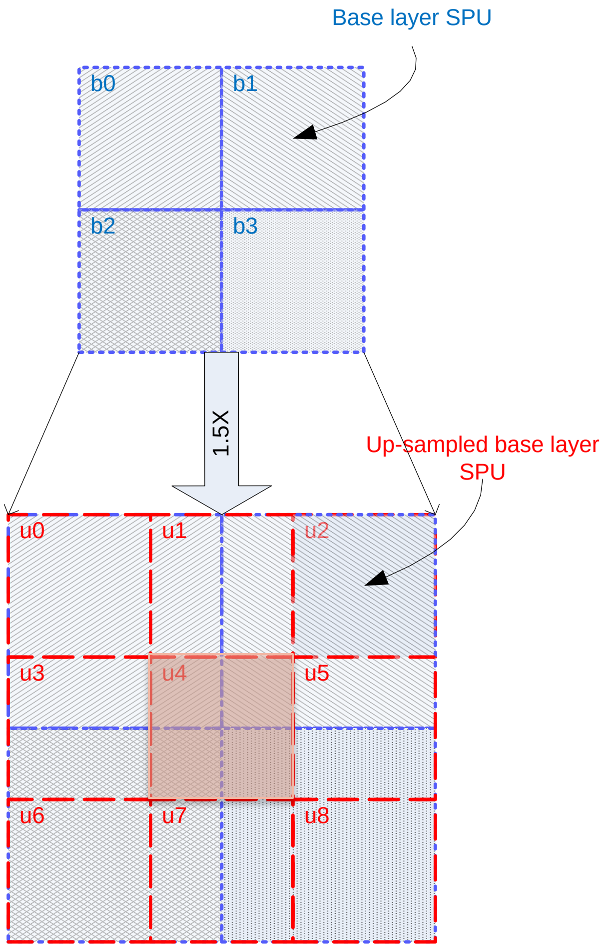
INTERDIGITAL.

Motion Field Mapping

- MFM "creates" the motion field for the inter layer reference pictures in the "virtual layer", just like upsampling "creates" the pixel values
- Inter layer motion prediction accomplished using TMVP in EL, without any block level changes
- Motion field mapping:
  - Reference index mapping: ref^{ILR}
    - majority rule, tie breaking based on temporal POC distance
  - Motion vector mapping:

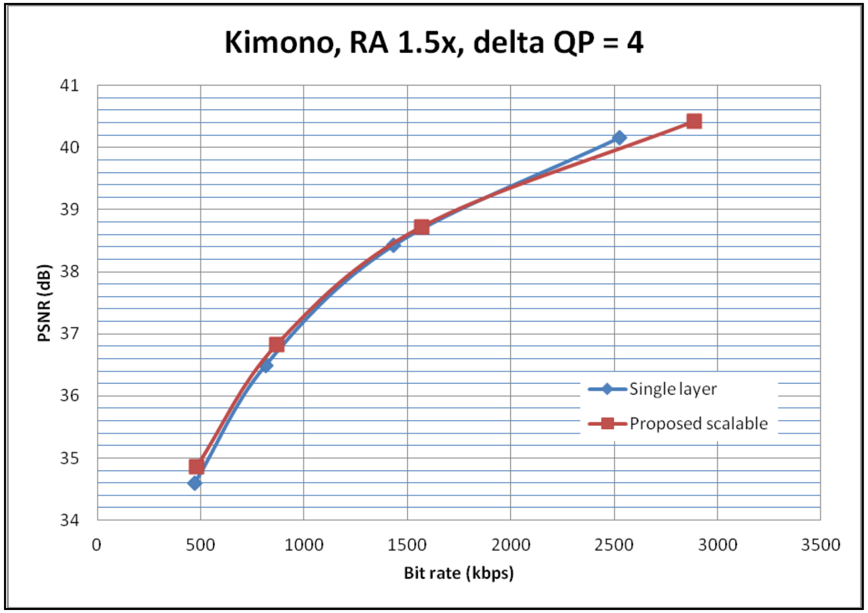
$$MV^{ILR} = N \frac{\sum_{i \in B} S_i \cdot MV_i^{BL}}{\sum_{i \in B} S_i} \qquad B = \{ j : ref_j = ref^{ILR} \}$$

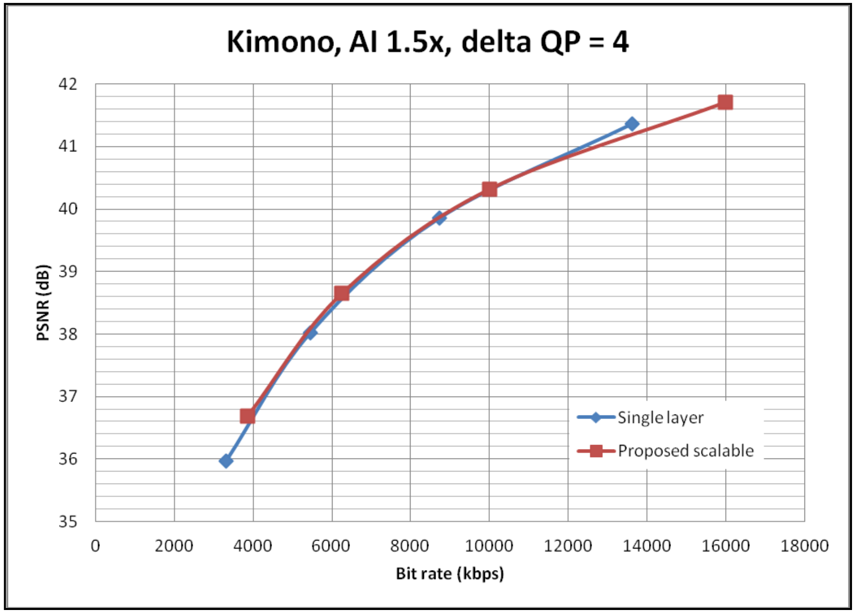
- Also provides motion field for EL skipped slice
- Our implementation uses the compressed MVs



EL Skipped Slice

- Sophisticated inter layer processing techniques can produce inter layer references with high quality
  - Esp. when base layer reconstruction quality is relatively high
- ELSkip skips CU level coding and directly copies from inter layer reference
  - Encoder makes decision based on RD
  - Currently signaled by a new NAL_UNIT_TYPE
  - Motion field is also copied from inter layer reference through MFM
  - Reduces decoding complexity
- In some cases, scalable coding efficiency comparable to single layer coding





INTERDIGITAL.

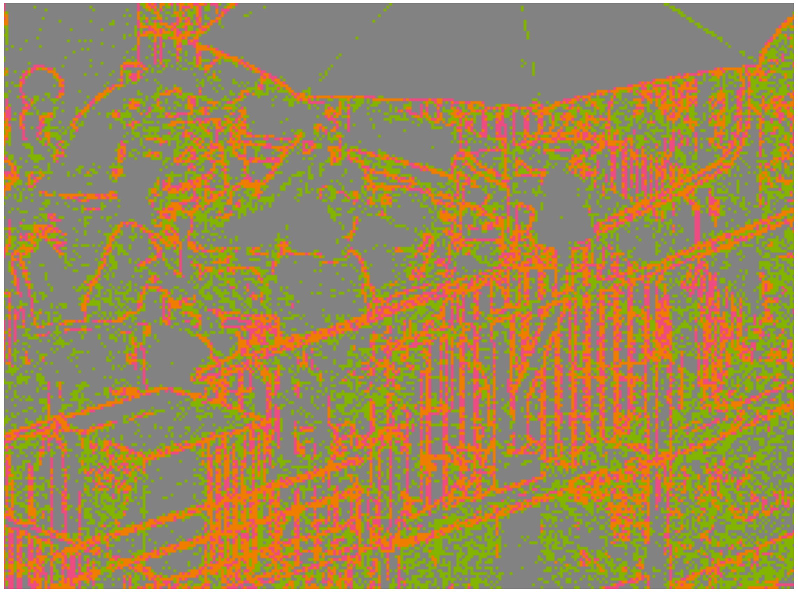
Adaptive Reference Picture Placement

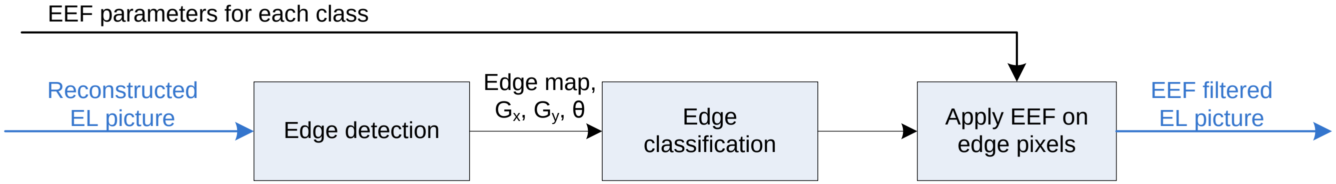
- Inter layer reference can offer better prediction than temporal reference pictures, e.g., when
  - Temporal distance is large
  - Base layer reconstruction is relatively high quality
- Adaptively decides inter layer reference picture placement in the reference picture lists
- In our implementation, the reference pictures are reordered using SATD based motion distortion

INTERDIGITAL.

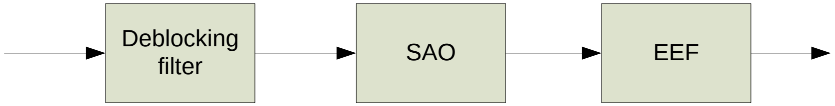
Edge Enhancement Filter

- Quantization tends to remove/distort high frequency components in the original video
  - Loss of detail, distorted edge/texture
- EEF is used to improve signal fidelity around edges
  - Canny edge detector
  - 16 edge pixel classes
  - Apply trained offset for each class



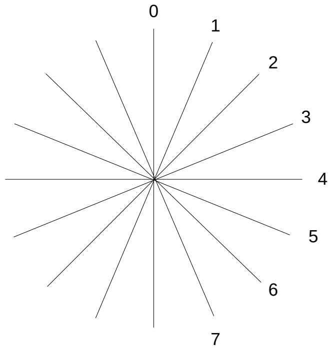


- EEF is applied after SAO in enhancement layer

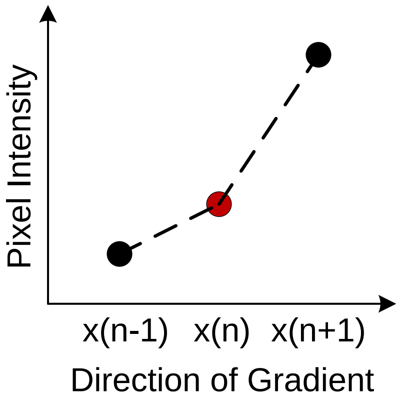


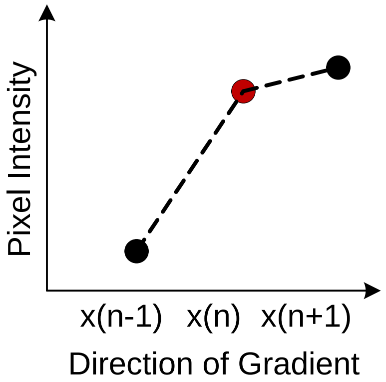
EEF: Edge Classification

1. Classify into 8 edge pixel groups based on edge direction θ



2. Within each group, classify into 2 sub-groups based on sign of 2nd order derivative in the edge gradient direction

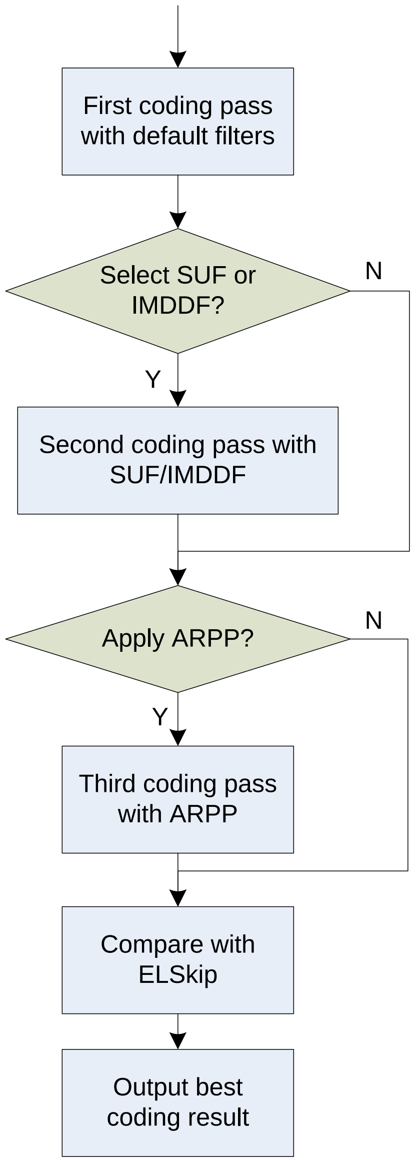




$$sign(n) = \begin{cases} -1 & if \;\; p(n) < (p(n-1) + p(n+1))/2 \\ +1 & if \;\; p(n) > (p(n-1) + p(n+1))/2 \end{cases}$$

Multi-pass encoding

- Picture level multi-pass encoding with early termination
  - Worst case 3 coding passes
  - Can be further consolidated by combining SUF/IMDDF pass with ARPP pass
- SUF/IMDDF early termination based on estimated R-D cost reduction
- ARPP early termination based on whether ILR picture placement is different from default



INTERDIGITAL.

Category 2: AVC base layer

- Extending the proposed framework to cover AVC base layer coding is straightforward
- Only pixel values from base layer reconstruction are used
- Not all proposed tools are applicable

Tool name	Applicability (yes/no)
SUF	Yes
IMDDF	No
MFM	No
ARPP	Yes
ELSkip	Yes
EEF	Yes

- Further gain to be expected when additional base layer information can be used

Syntax and Semantics

- Very simple syntax changes (high level syntax only)
  - nuh_layer_id_plus1
  - New NAL_unit_type for ELSkip slices
  - Adaptation Parameter Set
    - Very suitable for carrying picture level processing parameters
    - upsample_method_idc

Index	Name
0	DEFAULT
1	SUF
2	IMDDF

    - EEF parameters:16 offsets per color component per slice

INTERDIGITAL.

Performance Evaluation (EL-only actual rate)

	Category 1					Category 2	
	AI 2x	AI 1.5x	RA 2x	RA 1.5x	RA SNR	RA 2x	RA 1.5x
Y	-34.3%	-54.7%	-25.3%	-44.7%	-33.0%	-24.2%	-40.8%
U	-23.0%	-47.6%	-9.0%	-33.5%	-16.9%	-8.0%	-30.7%
V	-23.4%	-47.4%	-7.5%	-30.8%	-12.5%	-6.4%	-27.8%

Simulation Platforms

Linux grid for encoding

CPU	Intel Xeon E5-2650 / Clock speed 2.00GHz Intel Xeon X5675 / Clock speed 3.07GHz
Memory for each core	2 GB
HDD characteristics	-Hard Drive Capacity: 300GB -Interface: Serial Attached SCSI -External Data Transfer Rate: 6.0 GB/sec -Rotational Speed: 10,000 rpm
Compiler	GCC version 4.4.5
Operating System (OS)	64 bit Linux

Windows server for decoding

CPU	Intel Xeon X5675 / Clock speed 3.07GHz
Memory for each core	2 GB
HDD characteristics	-Hard Drive Capacity: 500GB -Interface: Serial Attached SCSI
Compiler	Visual studio 2010
Operating System (OS)	64 bit Windows Server 2003

Encoding Time and Memory Usage

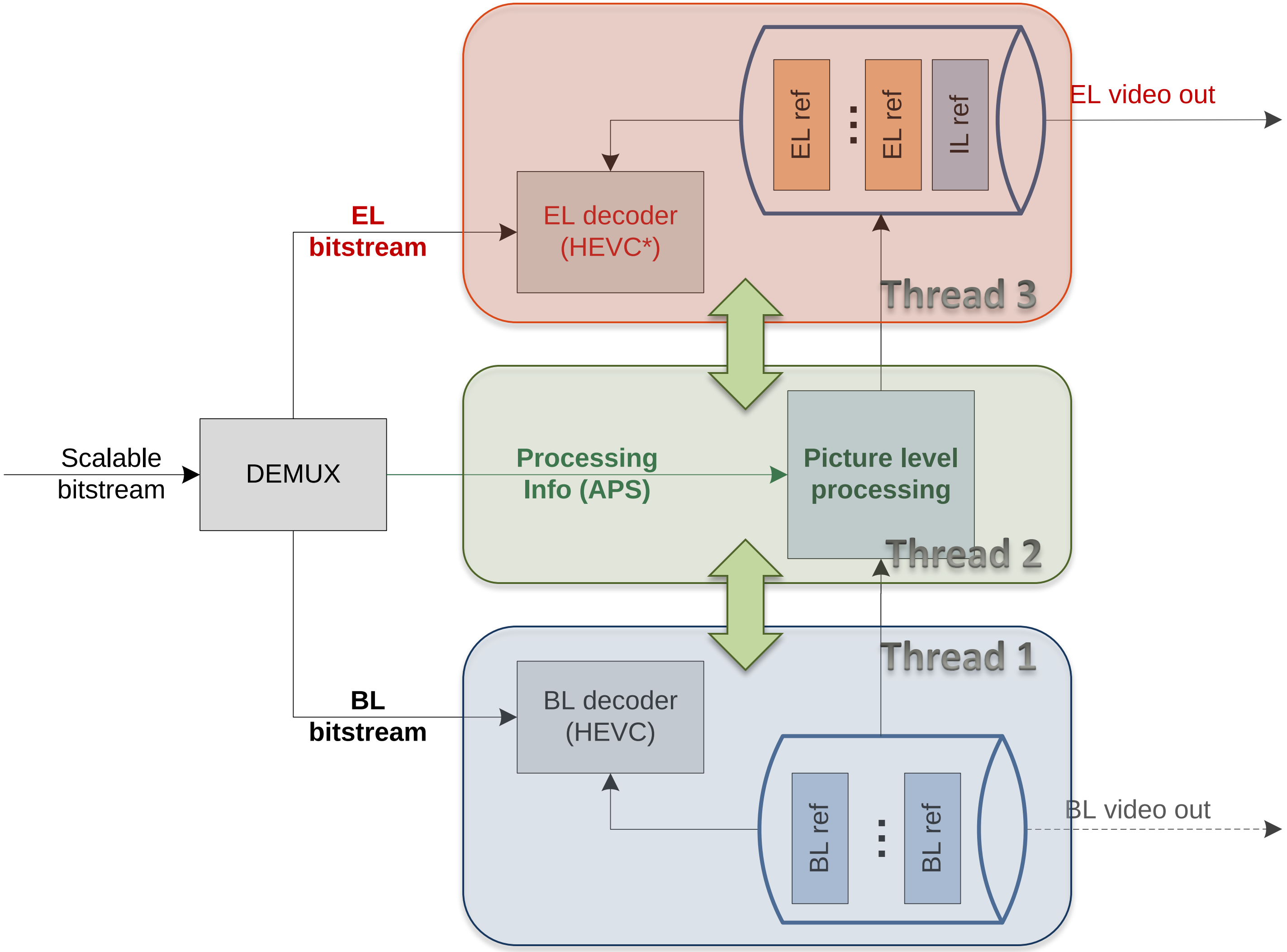
Config	Scalability	Class	Average encoding time (hours)	Average encoder memory usage
AI	2x	Class A	14.6	2.50 GB
		Class B	12.1	720 MB
	1.5x	Class B	8.7	750 MB
RA	2x	Class A	20.4	1.65 GB
		Class B	17.4	1.50 GB
	1.5x	Class B	16.4	1.68 GB
	SNR	Class A	27.4	3.50 GB
		Class B	16.7	2.00 GB

INTERDIGITAL.

Decoding Time and Memory Usage

Config	Scalability	Class	% decoding time increase relative to anchor	Average decoder memory usage
AI	2x	Class A	34%	2.08 GB
		Class B	118%	595 MB
	1.5x	Class B	58%	613 MB
RA	2x	Class A	164%	2.15 GB
		Class B	167%	610 MB
	1.5x	Class B	113%	630 MB
	SNR	Class A	73%	2.30 GB
		Class B	55%	635 MB

Implementation Aspects: Parallelization



Implemented a multi-threaded decoder of the proposed system

InterDigital

Closing Remarks

- MVC-like approach
  - Inter layer prediction accomplished by using inter layer reference as additional reference picture
  - No significant change to the enhancement layer codec
  - Aligned with 3D extensions in JCT-3V
- Low implementation complexity
- Designed for multi-threaded parallelization, especially at the decoder
  - Implemented a working prototype of multi-threaded decoder
- Question: for scalable extensions of HEVC, should we consider a low complexity setting that only requires high level syntax changes at enhancement layer?