JCTVC-K0041
Canon's response to the Joint CFP for Scalable Extension of HEVC

Fabrice Le Léannec
Sébastien Lasserre

JCTVC Meeting, Shanghai, 9 - 19 October, 2012

Content

- Canon Scalable Codec (CSC) overview
- Intra coding technology
- Inter coding technology
- Obtained performances
- Software considerations

CANON SCALABLE CODEC OVERVIEW

Canon Scalable Codec overview

- **Intra coding**:
  - New scalable coding technology providing **very low complexity**, enabling **spatial random access** and **high parallelism**
- Random access:
  - Enhancement Intra pictures processed with proposed low-complexity intra coding
  - Enhancement P/B pictures extend HEVC with **inter-layer prediction tools**
- CFP Categories addressed:
  - All intra with HEVC base layer
  - Random access with HEVC base layer
  - SNR

DESCRIPTION OF THE INTRA CODING TECHNOLOGY

Motivation

- Scalability usually increases complexity of non-scalable video codecs
- Here: scalability is used as a means to obtain a good trade-off between:
  - Low Complexity
  - Low Memory bandwidth
  - High coding efficiency
- Adapted for embedded light encoders

- Additional desired features:
  - Highly parallelizable coding/decoding processes
  - Spatial random access: ability to extract and decode a spatial area from the video stream

Canon

Principle of the Intra proposed approach

- Only **one coding mode used in the enhancement layer**
  - Inter-layer intra prediction (analogous to I_BL in H.264/SVC)
  - Suppression of all spatial dependencies between neighboring blocks in the enhancement pictures
    - **No spatial prediction**
    - **Context-free and non-adaptive entropy coding**

Block types (recall of JCTVC-I0190)

- A picture is segmented into **blocks**
- Each block belongs to a **block type** •

- A block type is defined by
  – its **geometry** (8x8, 16x16, 8x16, etc.)
  – **a label**
    - typically a label depending on the energy of the block

- Block type **syntax**
  – TU quad-tree
  – plus syntax for the label

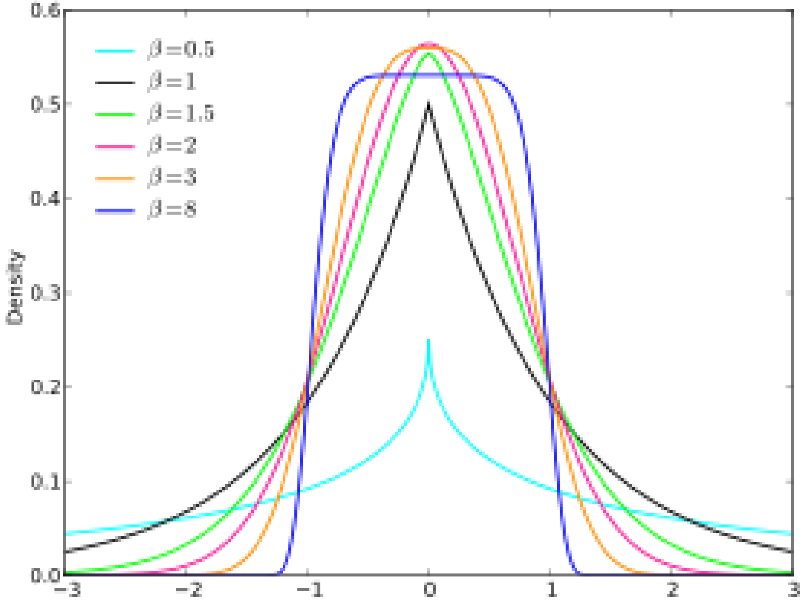
Canon

Modelling of DCT coefficients

- Probabilistic model: two-parameter **Generalized Gaussian distribution**

$$\boxed{\text{DCT } X \approx GGD(\alpha, \beta)}$$

- Depends on
  - **video content**
  - **coefficient index**
  - **block type**



exponent

$$GGD(\alpha, \beta, x) := \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp(-|x/\alpha|^{\beta})$$

$$\sigma^2 = \alpha^2 \Gamma(3/\beta) / \Gamma(1/\beta)$$

standard deviation

The encoding problem

$$\boxed{\text{minimize } R = \sum_{n} R_n(D_n) \quad \text{s.t.} \quad \sum_{n} D_n^2 = D_t^2}$$

the global minimization problem

total rate

rate associated to the encoding of the n-th DCT coeff

distortion due to the encoding of the n-th DCT coeff

total distortion target

- Splitting into two easier sub-problems
  - **find optimal quantizers on GGD channels**
  - **find the above minimization problem on optimal quantizers**
- Advantages
  - first sub-problem solved off-line once for all
  - second problem is tractable since one knows the rate-distortion curves $R_n(D_n)$ of the optimal quantizers

Canon

The theorem of equal merits for a block type

- **Merit of a DCT coefficient** is defined = basically R(D) encoding slope

- An **optimal encoding**, for a block type, is as below
  - All **encoded DCT coefficients** have the same merit after encoding
  - All **non-coded coefficients** have a merit smaller than the common merit of encoded coefficients

$$\boxed{\forall \text{coded DCT index } n \quad M_{\bullet} = M_{n,\bullet}}$$

- The optimal encoding **solution is unique**

- From this solution, one deduces **which optimal quantifiers** must be used for the encoding

Canon

Cascade of merits and choice of quantifiers

- Encoder side only
  - A **video merit** is provided by the user
    - by definition, this is the common merit of encoded DCT coefficients

$$\boxed{M_{Video}}$$

  - A **frame merit** is deduced for each frame
    - based on balance of PSNR-based quality balance between frames
    - signaled in the bit-stream

$$\boxed{M_{Video} := \frac{\Delta \mathrm{PSNR}}{\Delta R_F} = M_F / D_F^2}$$

- Encoder and decoder (synchronous)
  - A **block type merit** is deduced for each block type
    - based on the geometry of the blocks
  - A **DCT merit** is deduced for DCT coefficient
    - by using the theorem of equal merits
  - A **quantizer** is deduced for DCT coefficient
    - by calculating a DCT distortion target from the DCT merit

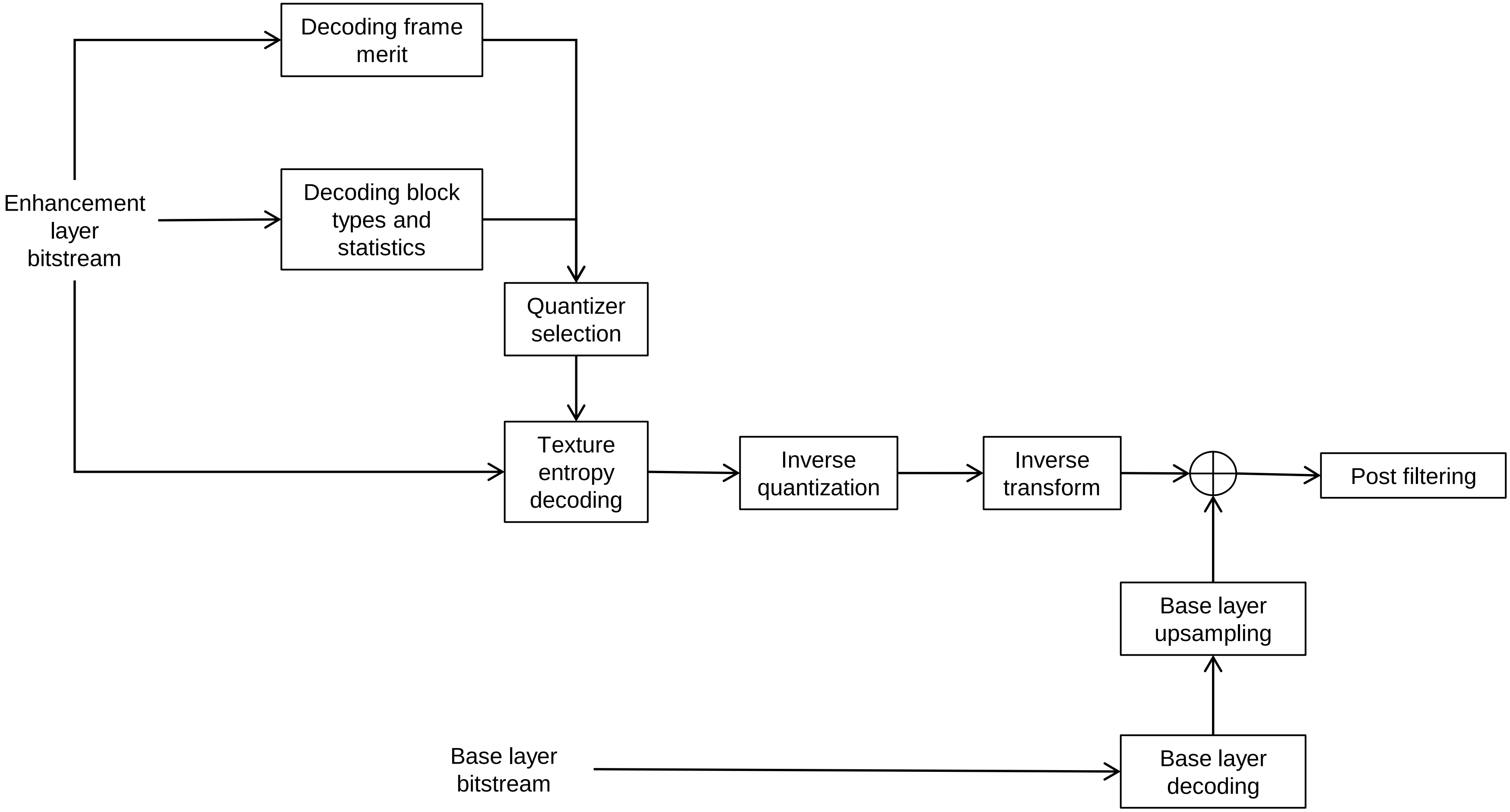
$$\boxed{M_{\bullet} = N_{\bullet} M_F}$$

$$\boxed{\forall \text{coded DCT index } n, \quad M_{\bullet} = M_{n,\bullet}}$$

$$\boxed{M_{\bullet,*} = \frac{2D_{n,\bullet,*}^2}{f'(-\ln(D_{n,\bullet,*} / \sigma_{n,\bullet,*}))}}$$

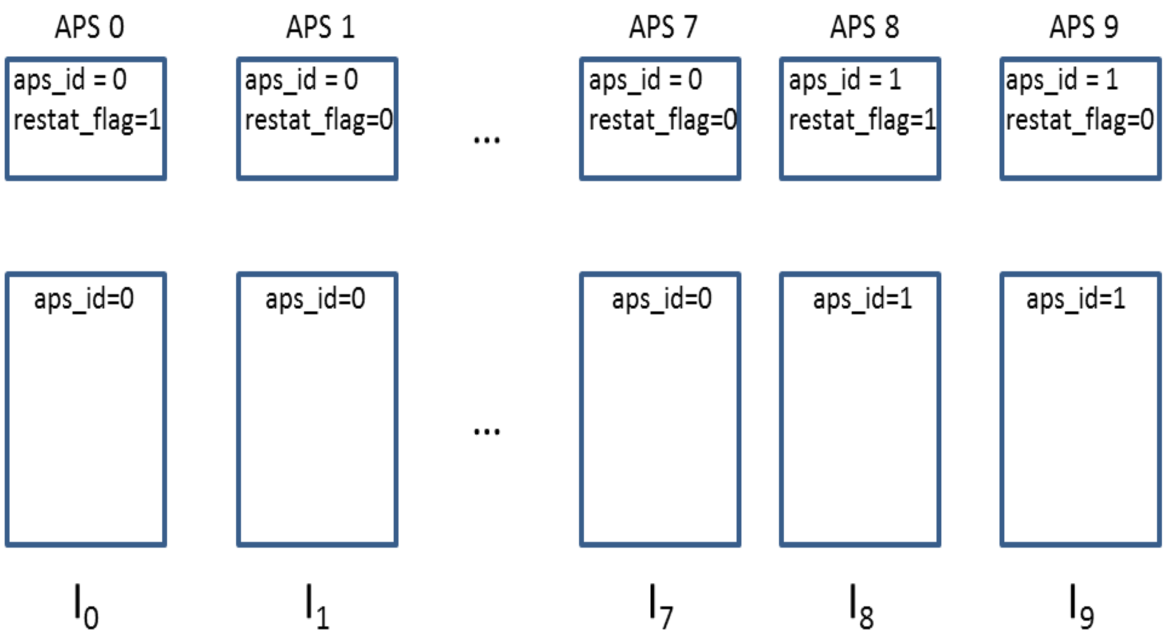
Canon

Intra decoder architecture



NAL unit containers for statistics

- Coding of GGD statistics spread over several successive frames
  - Coded in APS NAL units
  - “Restat” frames:
    - GGD stastistics coded from scratch
  - “Non-Restat” frames:
    - GGD statistics of previous frame are re-used
    - Additional statistics for new DCT channels coded in APS NAL unit

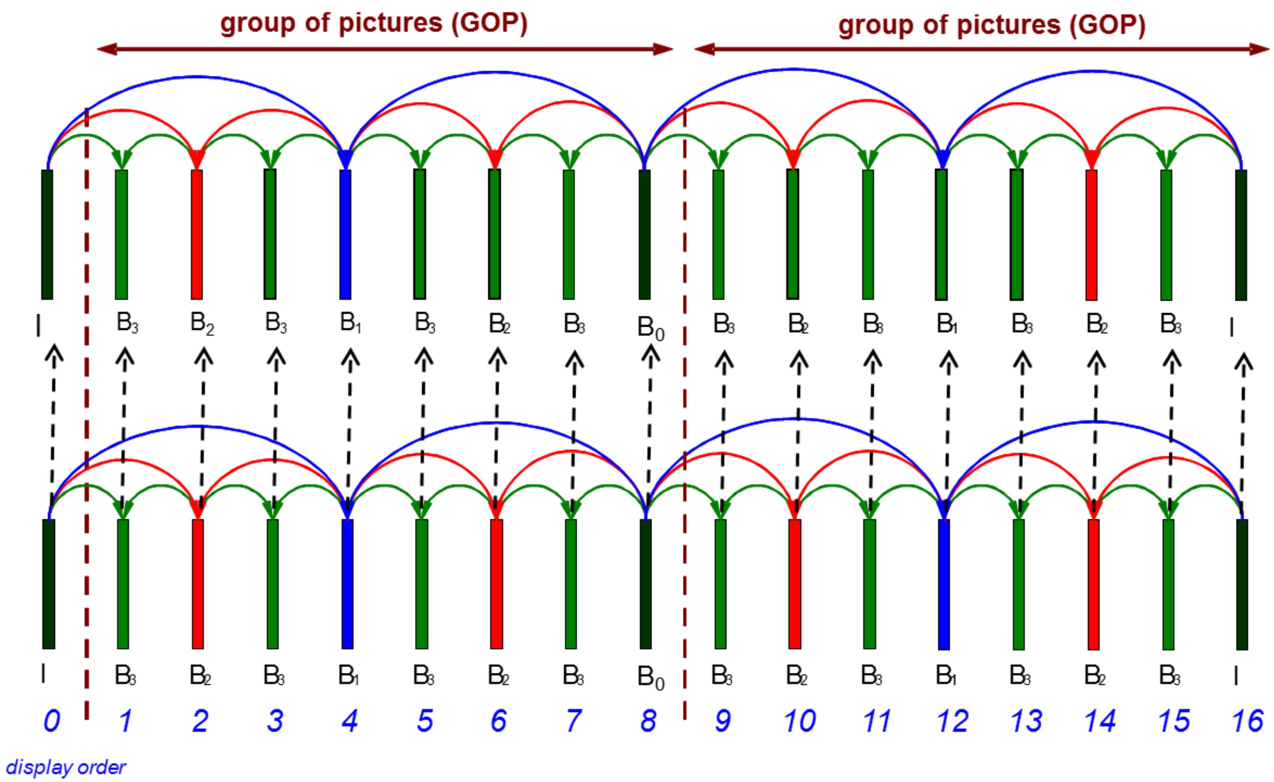


Canon

INTER CODING TECHNOLOGY

Random access coding structure

- Same coding structure in base and enhancement layer



- Intra pictures: coded with proposed low-complexity technology
- B pictures: HEVC-like coded with added inter-layer prediction tools

Inter-layer prediction tools

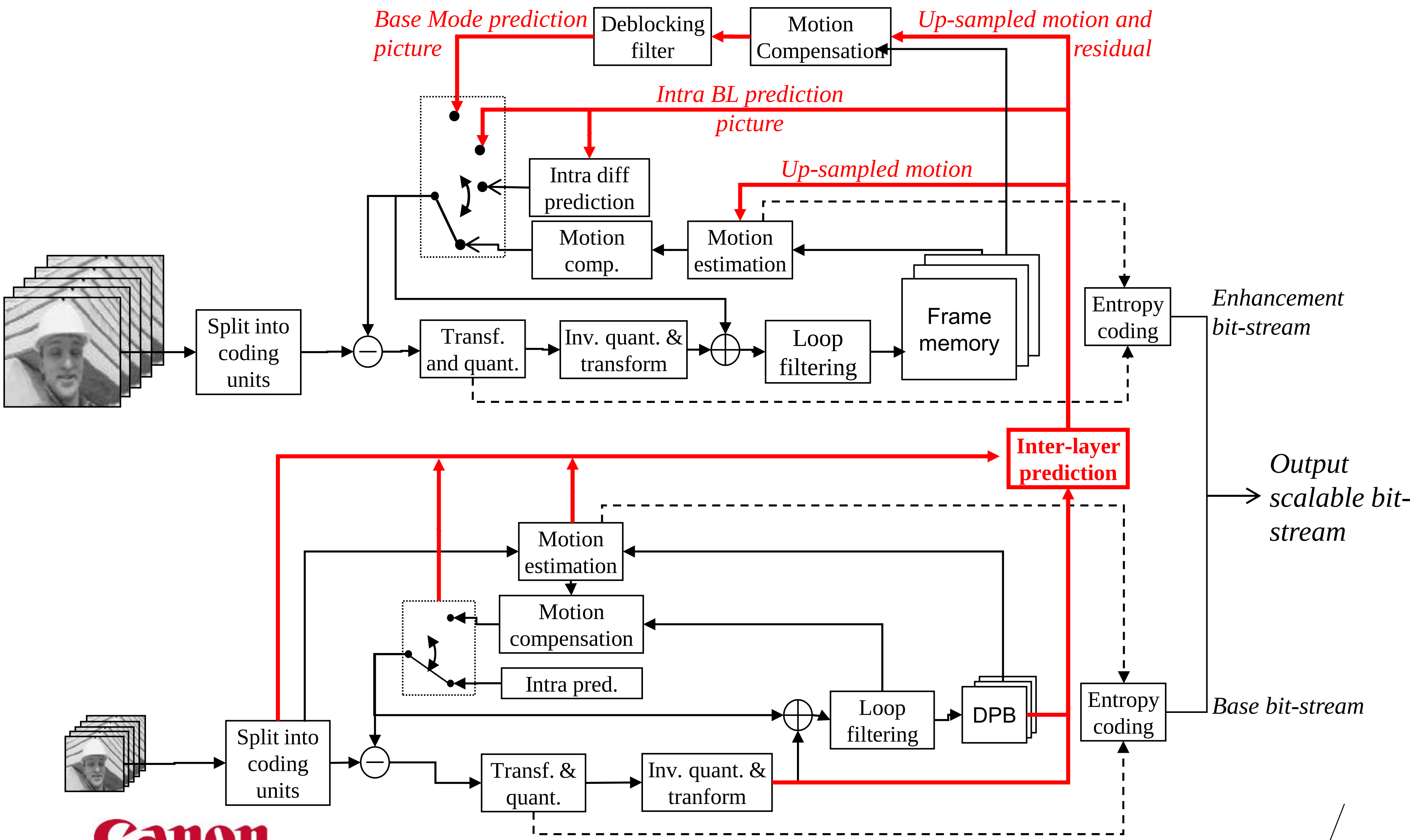
- Intra BL
- Base Mode
- Intra Diff (JCTVC-F290, replaces HEVC intra prediction)
- Generalized Residual Inter-Layer Prediction (GRILP)
- Motion Information Inter-Layer Prediction

Other tools

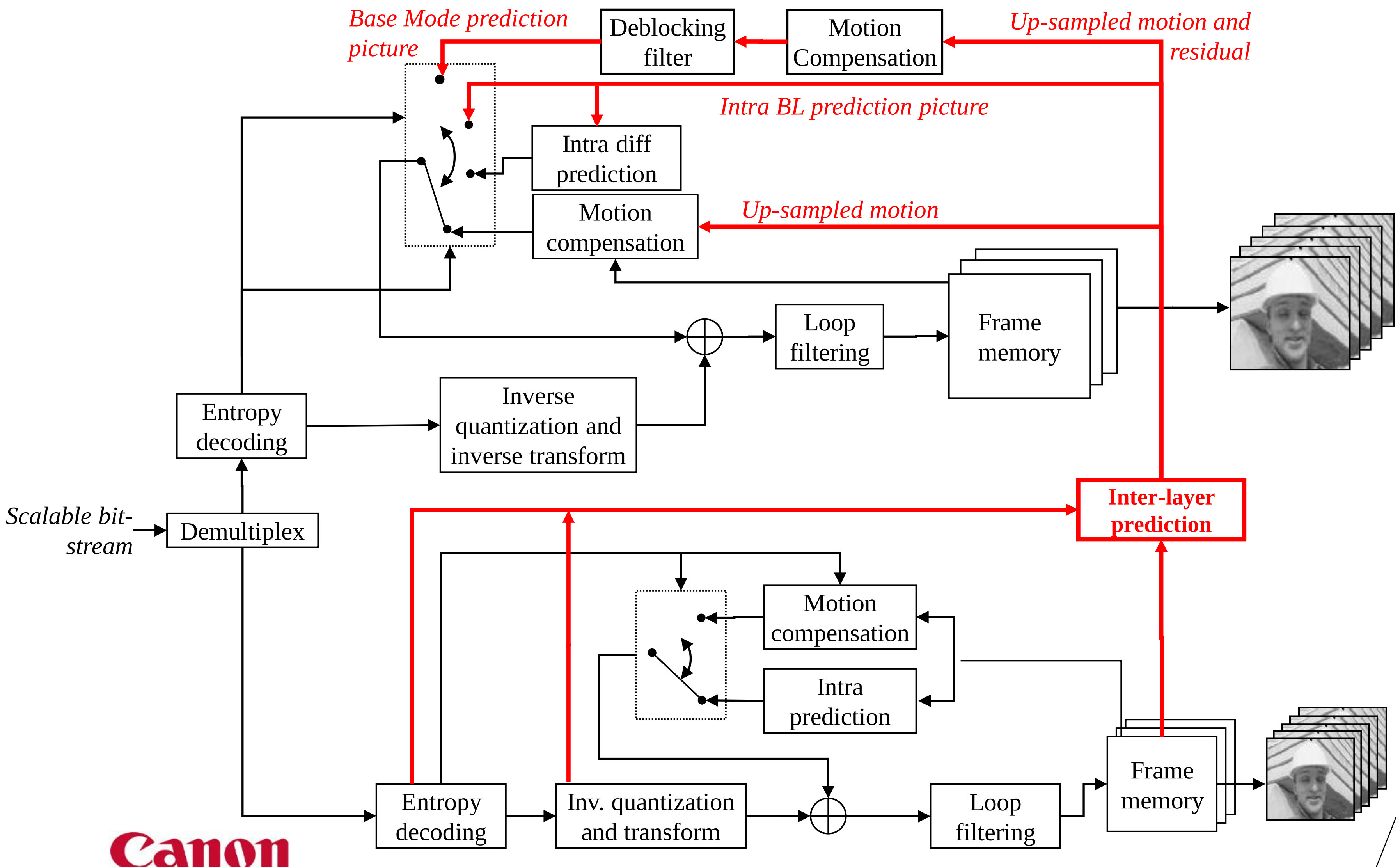
- In-loop filtering
  - DBF
  - SAO: HM6.1 picture level, improved JCTVC-G246, JCTVC-G290
  - ALF: HM6.1 picture level
- IBDI
- No AMP
- No NSQT
- No LM mode
- HM6.1 Inter modes

Canon

Encoder overview



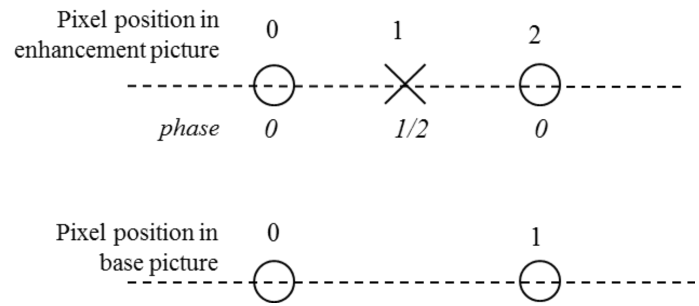
Decoder overview



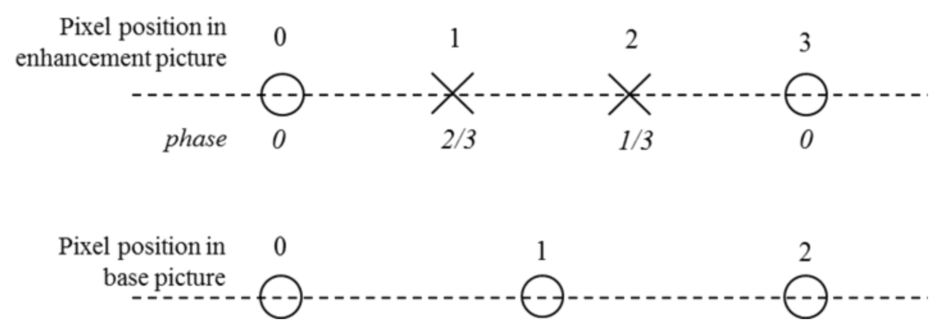
Inter-layer prediction tools: Intra BL

- Intra BL
  - Prediction of a Coding Unit from its co-located area in the up-sampled base picture
  - Up-sampling with DCT-IF interpolation filters

Phase	Luma taps	Chroma taps
0	{0, 0, 0, 64, 0, 0, 0, 0}	{ 0, 64, 0, 0}
1/3	{-1, 4, -11, 52, 26, -8, 3, -1}	{-5, 50, 22, -3}
½	{-1, 4, -11, 40, 40, -11, 4, -1}	{-4, 36, 36, -4}
2/3	{-1, 3, -8, 26, 52, -11, 4, -1}	{-3, 22, 50, -5}



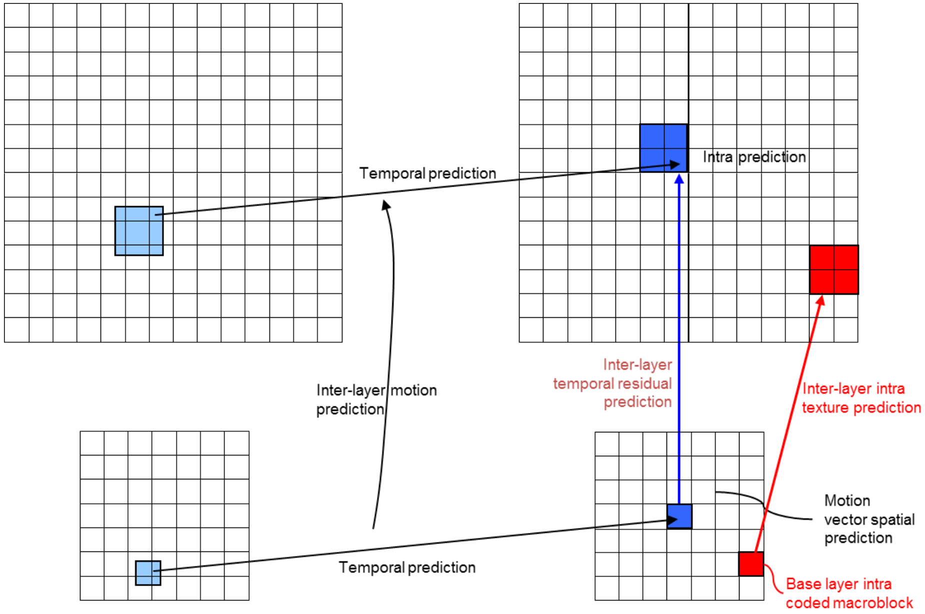
Ratio 2.0



Ratio 1.5

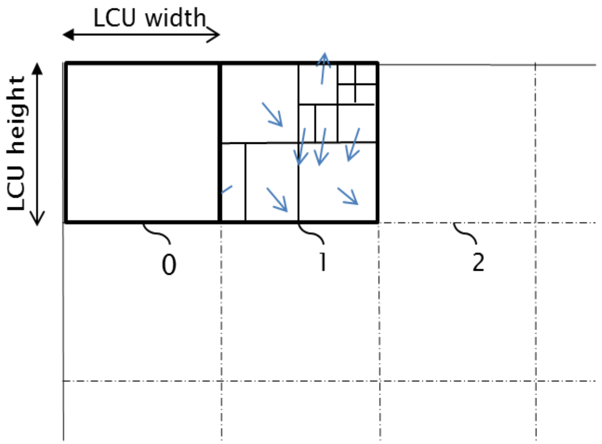
Inter-layer prediction tools: Base Mode

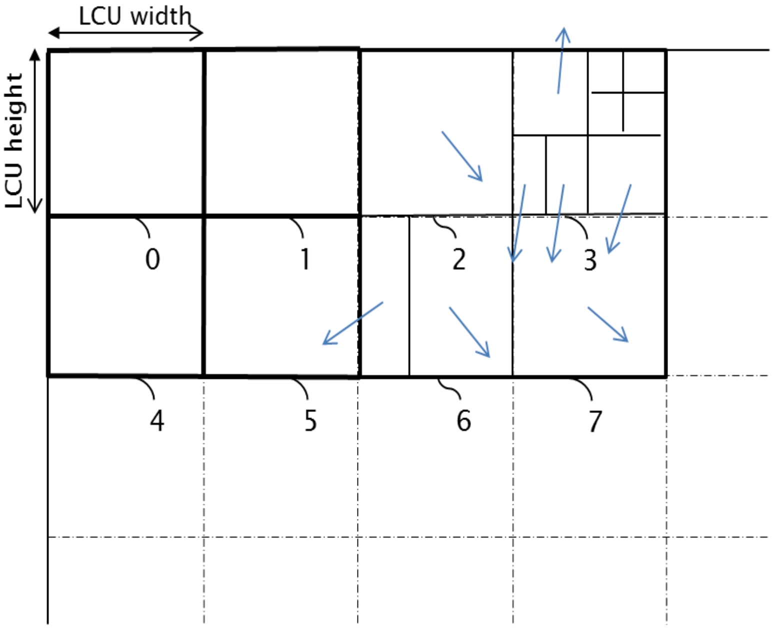
- Base Mode
  - Construction of a “Base Mode” prediction picture
    1. Inter-layer derivation of prediction information from the base layer
    2. Construction of “Base Mode” prediction picture:
       - Derived Intra CU ➔ Intra BL like prediction
       - Derived Inter CU
         - » MC prediction using derived prediction information from the base CU
         - » Inter-layer residual prediction with bi-linear interpolation filter
    3. Deblocking filtering of obtained “Base Mode” prediction picture
       - Boundaries of inter-layer derived CU’s and PU’s



Inter-layer derivation of prediction info

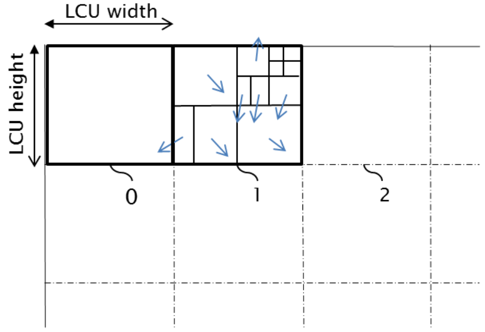
- Dyadic case:
  - Straightforward Derivation of CU quad-tree representation
  - High-level CU depth, PU partition, and Motion Vector derivation
  - Motion vector up-scaling
  - No merging between adjacent 4x4 blocks is needed

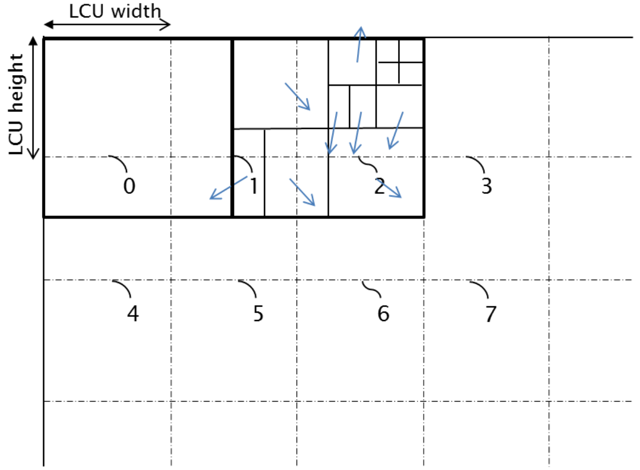




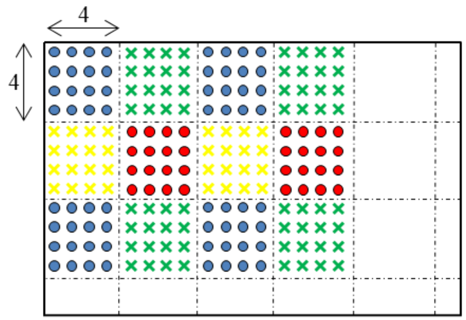
Inter-layer prediction info derivation

- Case of ratio 1.5
  - Derivation of CU representation is different from the dyadic case:

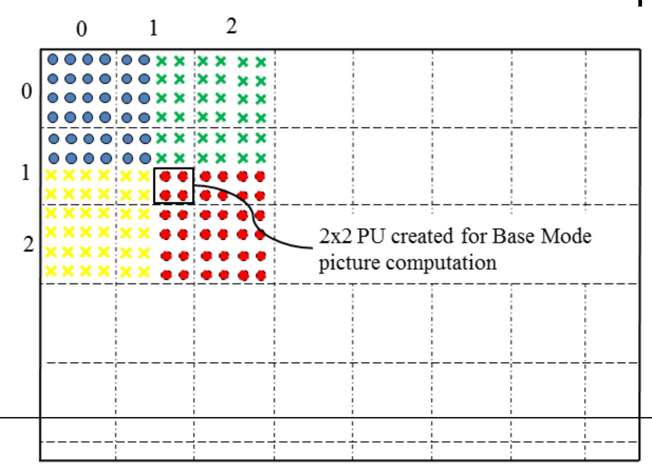




- Adopted approach for ratio 1.5:
  - Inter-layer prediction info derivation on a 4x4 block basis
  - Creation of temporary 2x2 CUs to preserve base motion info as much as possible

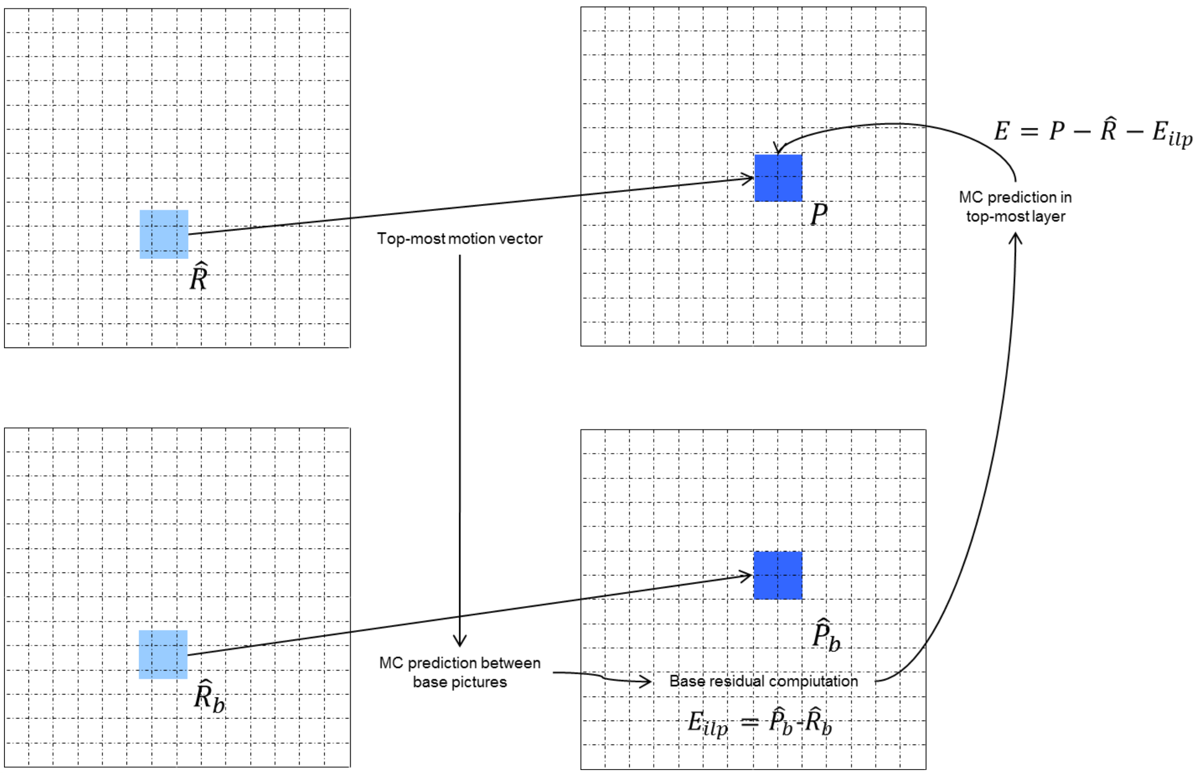


Base picture



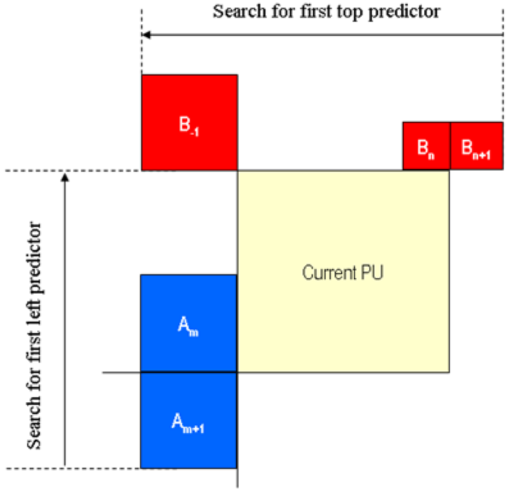
Generalized Residual Inter-Layer Prediction

- Inter-layer prediction of base layer temporal residual
- Base residual computed
  - Between fully reconstructed, up-sampled, base picture (multiple loop decoding)
  - According to enhancement motion vector
- Signaled on the CU level for Inter Coding Units



Inter-Layer Prediction of motion information

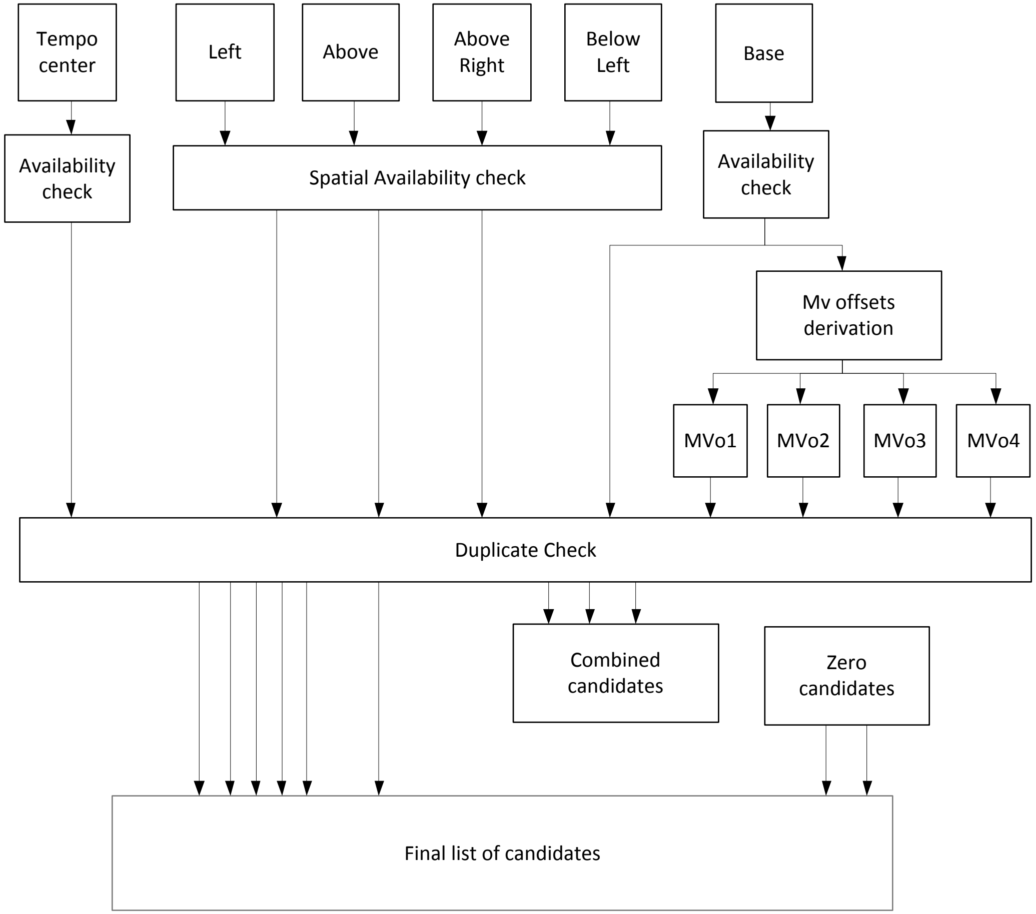
- AMVP prediction list increased from 2 to 3
  - First: temporal motion vector predictor
  - Second: spatial predictor
  - Third: co-located base motion vector
- Spatial prediction of motion information
  - Base Mode motion info added to the set of candidates
- No specific syntax for ILP of motion information



Canon

Modified Merge derivation

- Number of candidates in final list increased to 10
- Temporal predictor is first in the list, only centered position considered in EL
- Motion information memory compression is disabled
- Spatial candidates added to the list after temporal ones. Same spatial derivation as in base layer
- Co-located BL motion vector added to the end of the list
- 4 offset candidates derived from base motion vector candidate (4 and -4 on x and y)
- Duplicate check used on this list of 10 candidates, then combined candidates and zero candidates derivation applied if some position left available



PERFORMANCES

Summary of compression performances

- With PSNRs computed on 8-bit decoded sequences

"EL+BL actual rate"

Reference: (**Base+Simul. HR Ref.** bit-rate, **Simul. HR Ref.** PSNR)

Tested: (**Base+Scalable Enhan Tested** bit-rate **Scalable Enhan Tested** PSNR)

	All Intra HEVC 2x			**All Intra HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-27,1%	-28,1%	-28,4%			
Class B	-22,0%	-17,0%	-17,4%	-33,6%	-30,1%	-29,4%
Overall	-24,6%	-22,6%	-22,9%	-33,6%	-30,1%	-29,4%
	-24,6%	-22,5%	-23,1%	-33,6%	-33,1%	-32,6%
	Random Access HEVC 2x			**Random Access HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-26,2%	-14,6%	-13,7%			
Class B	-23,5%	-14,6%	-12,1%	-32,8%	-21,9%	-18,3%
Overall	-24,9%	-14,6%	-12,9%	-32,8%	-21,9%	-18,3%
	-24,9%	-14,6%	-12,9%	-32,8%	-21,8%	-18,3%
	Random Access HEVC SNR			All intra HEVC SNR		
	Y	U	V	Y	U	V
Class A+	-30,3%	-21,6%	-20,5%	-32,0%	-34,7%	-35,3%
Class B	-28,1%	-20,0%	-17,5%	-30,8%	-34,0%	-33,5%
Overall	-29,2%	-20,8%	-19,0%	-31,4%	-34,3%	-34,4%
	-29,2%	-20,8%	-19,0%	-31,4%	-34,5%	-34,6%

Canon

Summary of compression performances

- With PSNRs computed on 8-bit decoded sequences

"EL+BL target rate"
Reference: (**Base+Simul. HR Ref.** bit-rate, **Simul. HR Ref.** PSNR)
Tested: (**Base+Target** bit-rate, **Scalable Enhan Tested** PSNR)

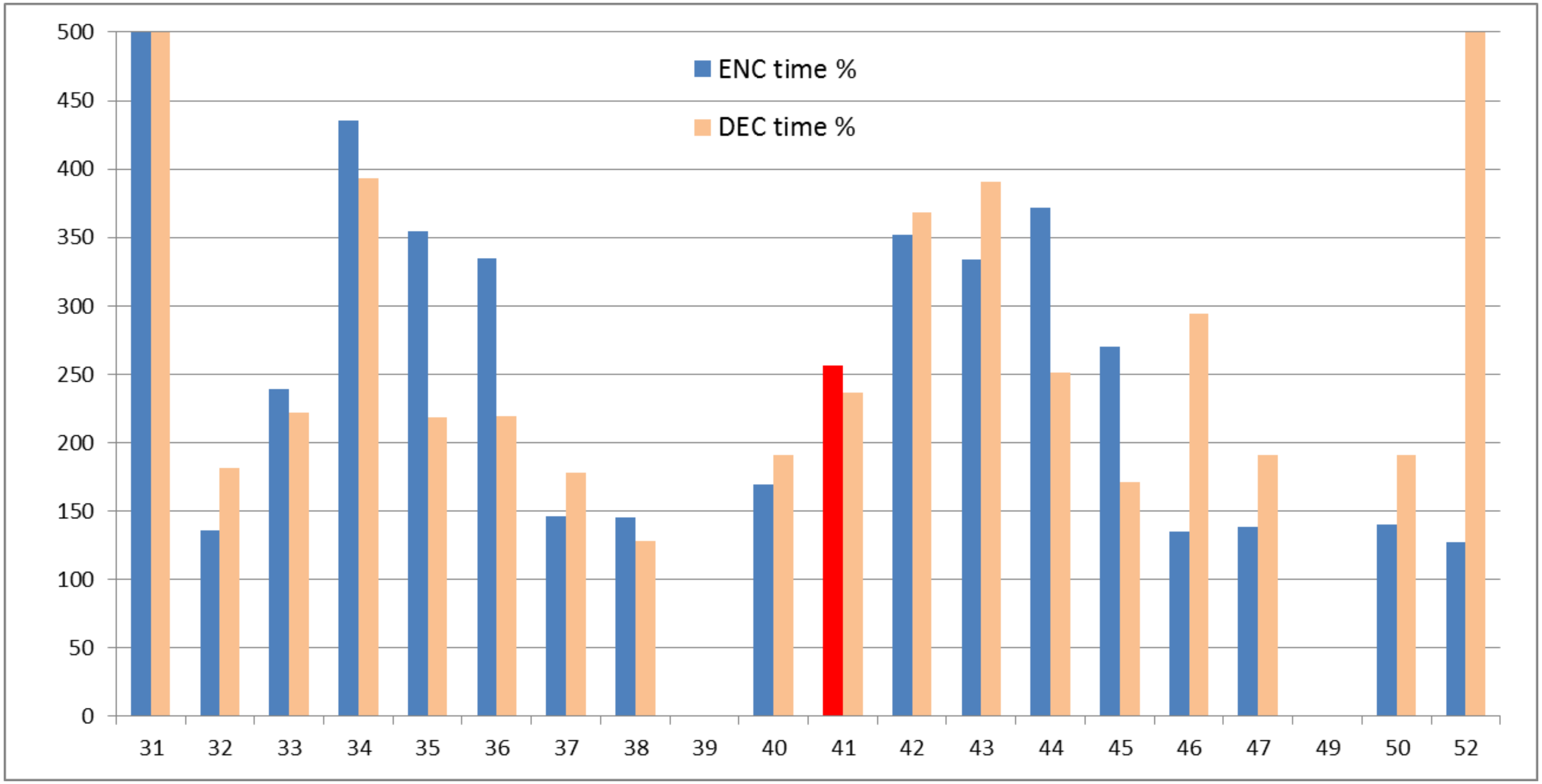
	All Intra HEVC 2x			**All Intra HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-26,0%	-27,0%	-27,2%			
Class B	-20,3%	-15,2%	-15,5%	-32,8%	-29,2%	-28,5%
Overall	-23,2%	-21,1%	-21,4%	-32,8%	-29,2%	-28,5%
	-23,2%	-21,0%	-21,6%	-32,8%	-32,3%	-31,8%

	Random Access HEVC 2x			**Random Access HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-26,0%	-14,4%	-13,5%			
Class B	-23,0%	-14,1%	-11,5%	-32,5%	-21,6%	-17,9%
Overall	-24,5%	-14,2%	-12,5%	-32,5%	-21,6%	-17,9%
	-24,6%	-14,2%	-12,5%	-32,6%	-21,5%	-17,9%

	Random Access HEVC SNR			All intra HEVC SNR		
	Y	U	V	Y	U	V
Class A+	-29,5%	-20,7%	-19,6%	-31,0%	-33,7%	-34,4%
Class B	-27,4%	-19,1%	-16,6%	-29,8%	-33,0%	-32,5%
Overall	-28,4%	-19,9%	-18,1%	-30,4%	-33,4%	-33,4%
	-28,4%	-19,9%	-18,1%	-30,4%	-33,5%	-33,6%

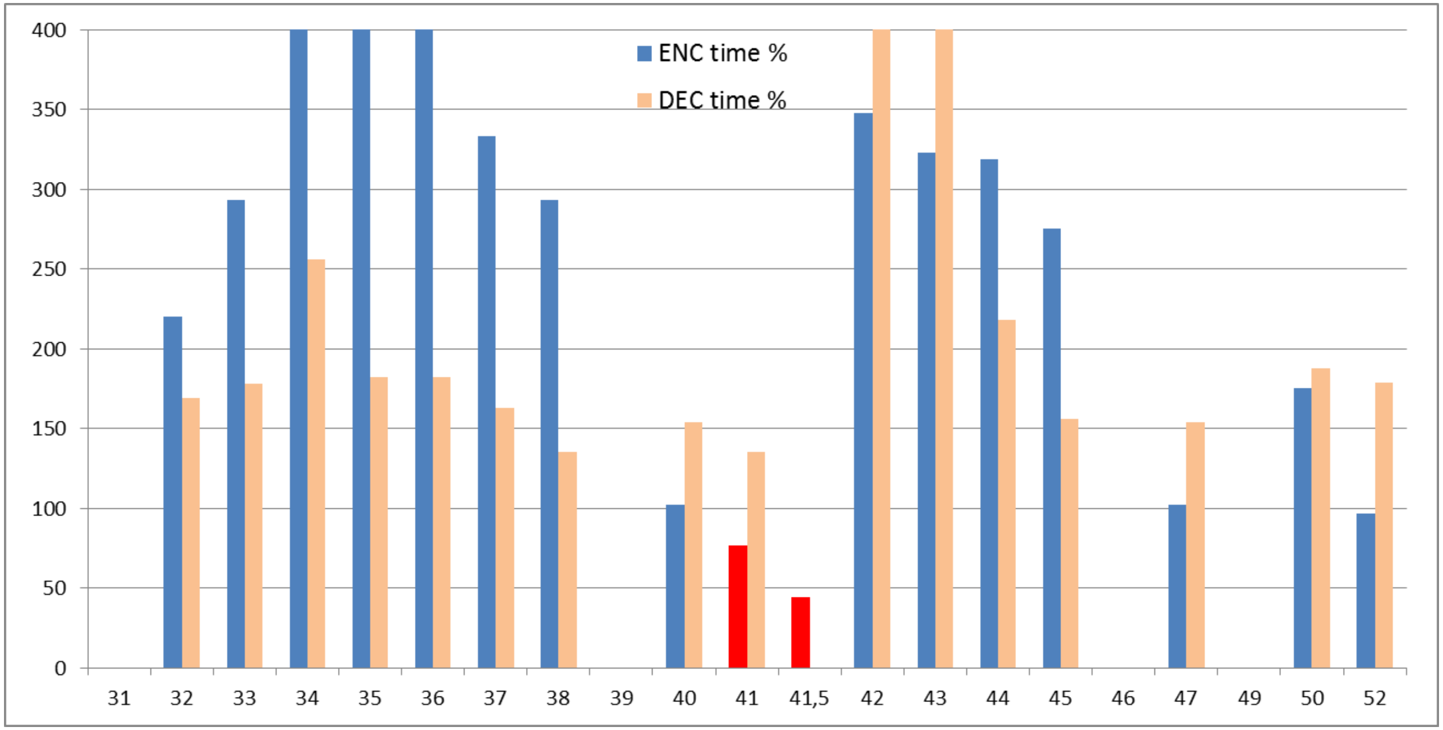
RA20 encoding and decoding times

- Comparative times obtained with QP 22 in enhancement layer



All Intra encoding and decoding times

- Comparative times obtained with QP 22 enhancement layer (AI20)



- Note: ALF takes more than 40 % of the All intra coding time !
- Without ALF (41.5): coding time = 44% (AI20) and 59% (AI15) of anchor
- Enhancement coding time ~ between 20% and 50% of the base layer coding time

More recent compression results

- With PSNRs computed on 8-bit decoded sequences
- AMP
- Improved Base Mode
- Imrpoved Inter-Layer motion derivation

"EL+BL actual rate"

Reference:	(**Base+Simul. HR Ref.** bit-rate,	**Simul. HR Ref.** PSNR)
Tested:	(**Base+Scalable Enhan Tested** bit-rate	**Scalable Enhan Tested** PSNR)

	All Intra HEVC 2x			**All Intra HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-27,1%	-28,1%	-28,4%			
Class B	-22,0%	-17,0%	-17,4%	-33,6%	-30,1%	-29,4%
Overall	-24,6%	-22,6%	-22,9%	-33,6%	-30,1%	-29,4%
	-24,6%	-22,5%	-23,1%	-33,6%	-33,1%	-32,6%

	Random Access HEVC 2x			**Random Access HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-27,5%	-13,4%	-4,1%			
Class B	-24,0%	-15,3%	-12,6%	-33,3%	-22,5%	-18,9%
Overall	-25,7%	-14,4%	-8,4%	-33,3%	-22,5%	-18,9%
	-25,7%	-14,3%	-8,4%	-33,4%	-22,4%	-18,6%

	Random Access HEVC SNR			All intra HEVC SNR		
	Y	U	V	Y	U	V
Class A+	-30,7%	-22,2%	-21,1%	-32,0%	-34,7%	-35,3%
Class B	-28,6%	-20,9%	-18,4%	-30,8%	-34,0%	-33,5%
Overall	-29,7%	-21,6%	-19,7%	-31,4%	-34,3%	-34,4%
	-29,7%	-21,5%	-19,7%	-31,4%	-34,5%	-34,6%

Canon

SOFTWARE CONSIDERATIONS

Software considerations

- Full test set passed for RA20, RA15, RA-SNR, AI20, AI15 with HEVC base
- Every provided bit-stream is correctly decoded
- No bug known so far in the source code
- We are ready to provide our source code as a basis for reference software

CONCLUSION AND PERSPECTIVE

Conclusion

- All Intra characteristics:
  - Low complexity Intra scalable coding**: complexity of single layer coding is reduced in spatial scalability**
    - **One single coding mode used**
    - Obtained coding efficiency shows limited cost of scalability compared to non-scalable HEVC
    - Coding/decoding process is highly parallelizable
    - Spatial random access with fine granularity can easily be provided
  - **Competitive rate-distortion performances obtained with a very low complexity INTRA scalable coding process**
- Random Access discussion:
  - **Set of inter-layer prediction tools added to standard HEVC**
  - **Good coding efficiency performance obtained (~3/4th proposal in terms of BD-rate)**
  - **Good synergy between B pictures and proposed scalable Intra coding system**
- We thank Vidyo for providing initial source code based on HM6.1

Summary of compression performances

- With PSNRs computed on 10-bit decoded sequences

"EL+BL actual rate"
Reference: (**Base+Simul. HR Ref.** bit-rate, **Simul. HR Ref.** PSNR)
Tested: (**Base+Scalable Enhan Tested** bit-rate **Scalable Enhan Tested** PSNR)

	All Intra HEVC 2x			**All Intra HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-27,8%	-29,7%	-30,4%			
Class B	-22,9%	-19,0%	-19,8%	-34,3%	-31,6%	-31,4%
Overall	-25,3%	-24,3%	-25,1%	-34,3%	-31,6%	-31,4%
	-25,4%	-24,3%	-25,2%	-34,3%	-34,6%	-34,5%

	Random Access HEVC 2x			**Random Access HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-27,0%	-17,9%	-18,1%			
Class B	-24,6%	-18,8%	-17,5%	-33,7%	-25,6%	-23,1%
Overall	-25,8%	-18,3%	-17,8%	-33,7%	-25,6%	-23,1%
	-25,8%	-18,3%	-17,8%	-33,8%	-25,4%	-22,8%

	Random Access HEVC SNR			All intra HEVC SNR		
	Y	U	V	Y	U	V
Class A+	-31,2%	-25,2%	-25,4%	-32,8%	-36,5%	-37,6%
Class B	-29,4%	-24,8%	-23,8%	-31,8%	-36,1%	-36,2%
Overall	-30,3%	-25,0%	-24,6%	-32,3%	-36,3%	-36,9%
	-30,3%	-25,0%	-24,6%	-32,3%	-36,5%	-37,2%

Summary of compression performances

- With PSNRs computed on 10-bit decoded sequences

"EL+BL target rate"
Reference: (**Base+Simul. HR Ref.** bit-rate, **Simul. HR Ref.** PSNR)
Tested: (**Base+Target** bit-rate, **Scalable Enhan Tested** PSNR)

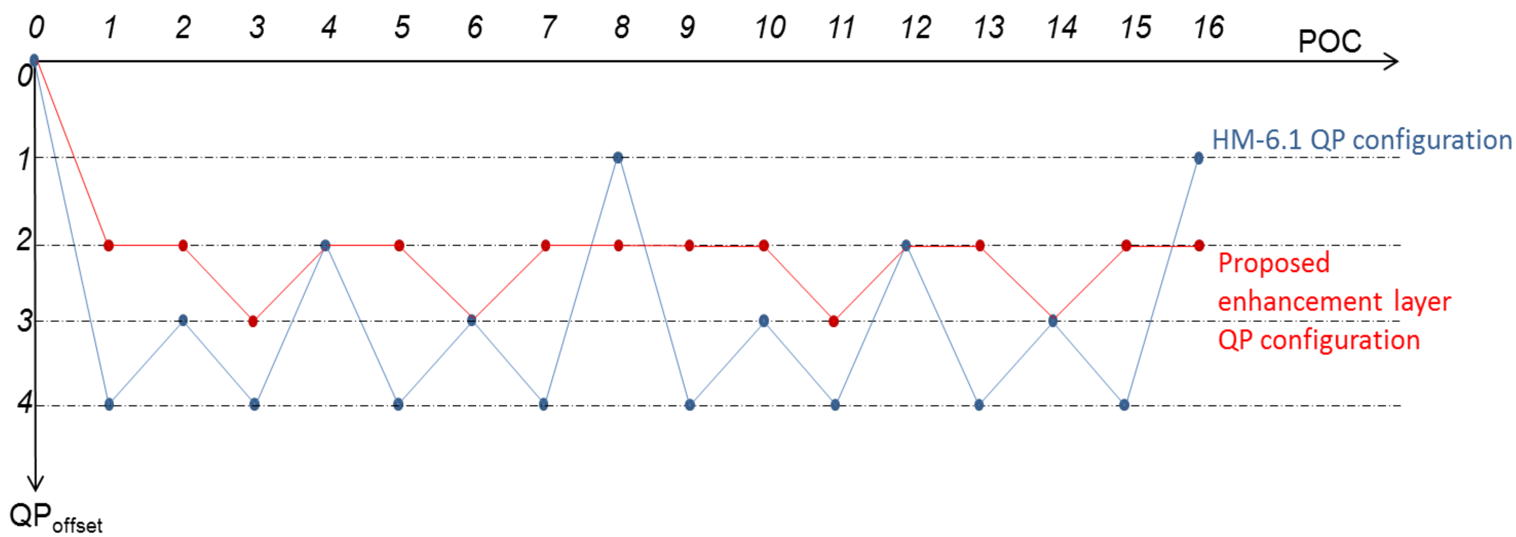
	All Intra HEVC 2x			**All Intra HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-26,7%	-28,6%	-29,3%			
Class B	-21,2%	-17,2%	-18,0%	-33,5%	-30,8%	-30,5%
Overall	-23,9%	-22,9%	-23,7%	-33,5%	-30,8%	-30,5%
	-24,0%	-22,8%	-23,8%	-33,5%	-33,8%	-33,7%

	Random Access HEVC 2x			**Random Access HEVC 1.5x**		
	Y	U	V	Y	U	V
Class A+	-26,8%	-17,7%	-17,9%			
Class B	-24,1%	-18,3%	-16,9%	-33,5%	-25,3%	-22,7%
Overall	-25,4%	-18,0%	-17,4%	-33,5%	-25,3%	-22,7%
	-25,5%	-18,0%	-17,4%	-33,5%	-25,1%	-22,5%

	Random Access HEVC SNR			All intra HEVC SNR		
	Y	U	V	Y	U	V
Class A+	-30,4%	-24,3%	-24,6%	-31,8%	-35,6%	-36,7%
Class B	-28,7%	-24,0%	-23,0%	-30,8%	-35,2%	-35,3%
Overall	-29,5%	-24,1%	-23,8%	-31,3%	-35,4%	-36,0%
	-29,5%	-24,1%	-23,8%	-31,3%	-35,5%	-36,3%

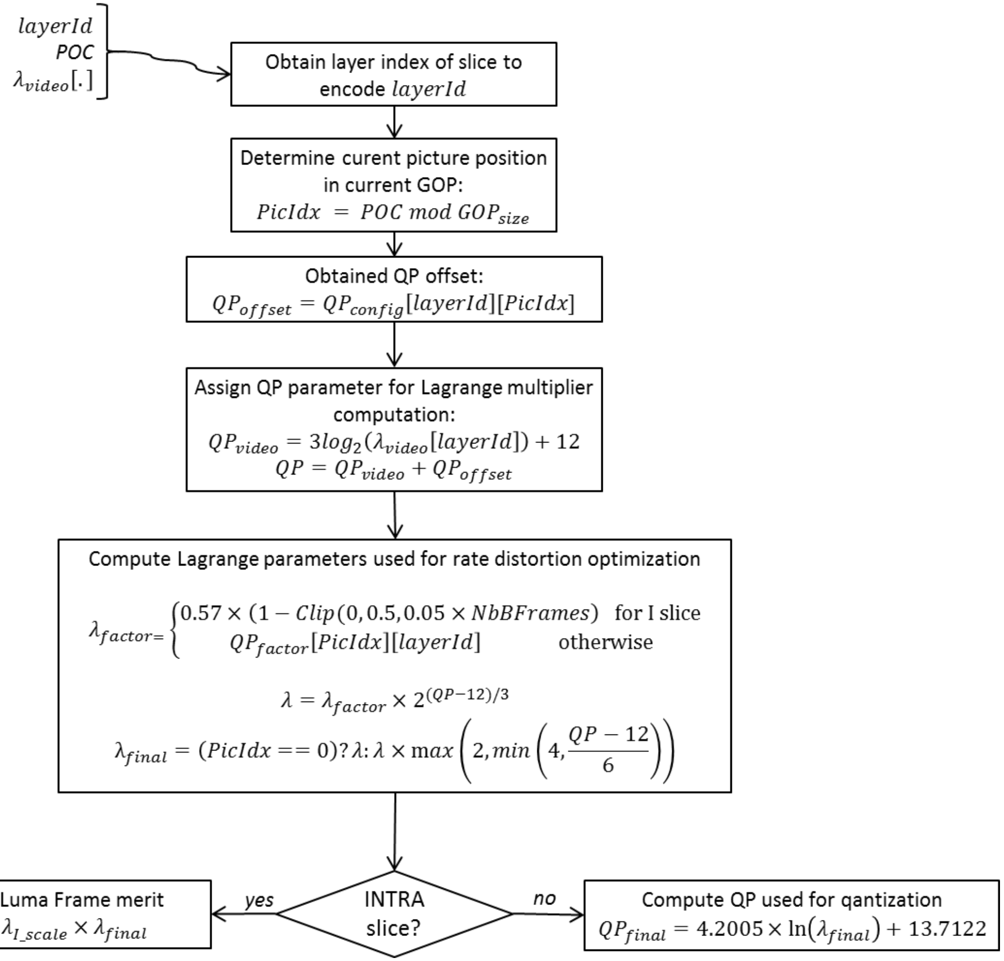
Encoder control method used for Random Access

- QP configuration used in the enhancement layer

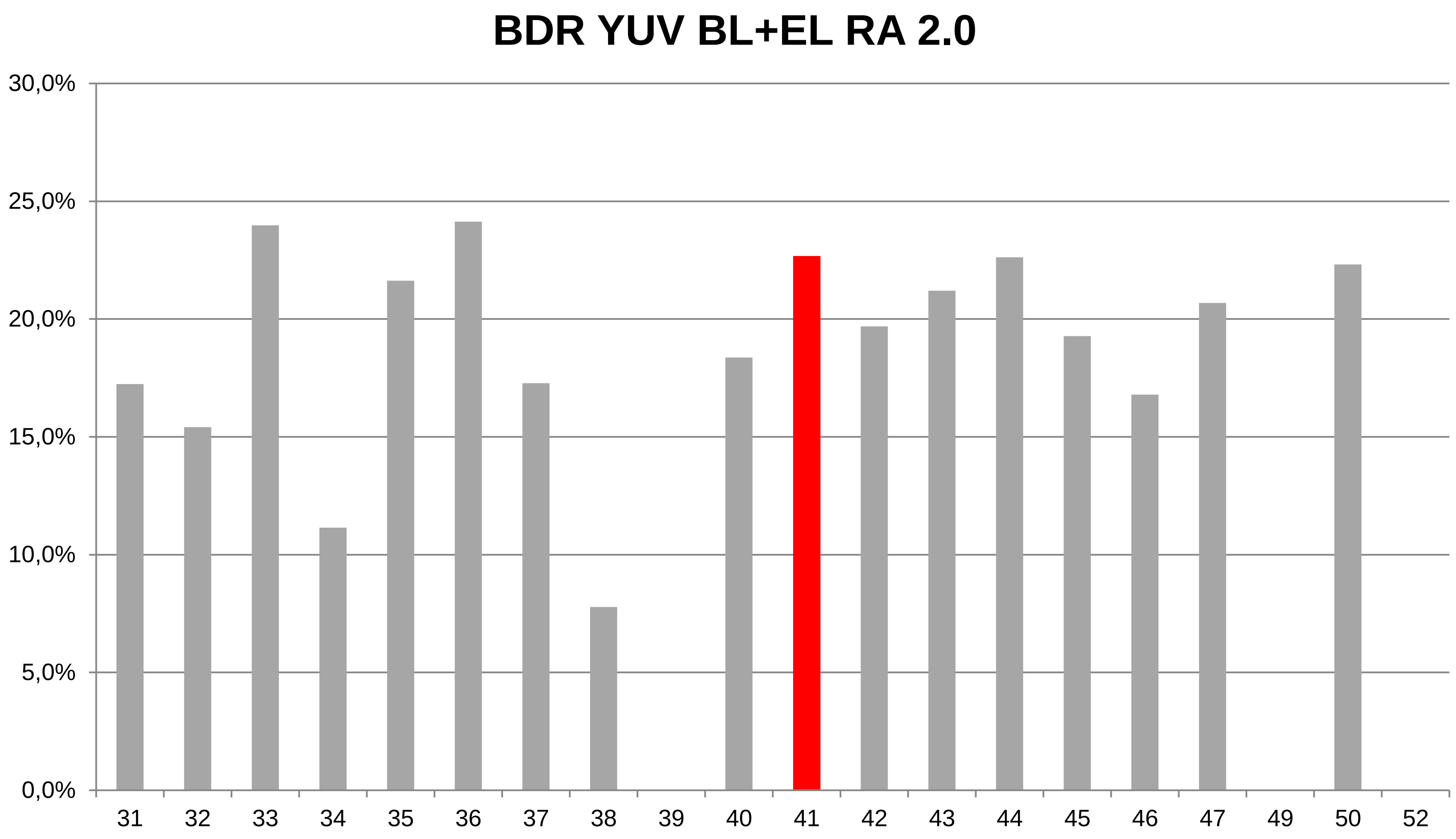


Lagrange parameter and QP used in RA

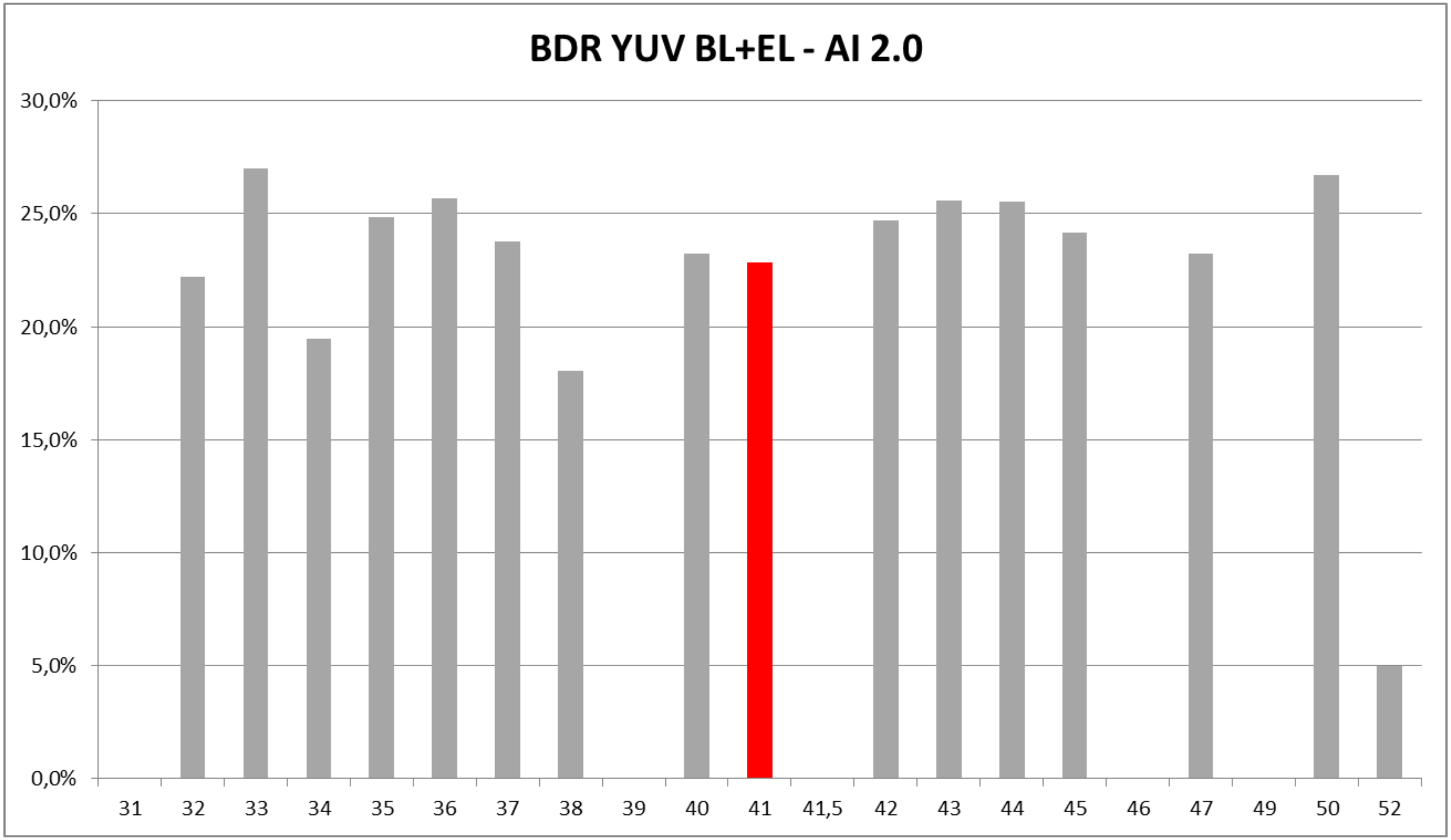
- Input: user param for enhancement layer $\lambda_{video}[layerId]$ parameter (video level)
- Output: QP_{final} and λ_{final} used to encode on the picture level



BD-Rate comparisons in RA20 Configuration



BD-Rate comparisons in AI20 Configuration



Canon